Diary No 738(2)
Date 1.2.74
J. N. Lib
Bureau Edul Psy: Rese

Journal of Experimental Psychology

DAVID A. GRANT, Editor
University of Wisconsin-Madison

HAROLD W. HAKE, Associate Editor
University of Illinois, Urbana

LEONARD E. ROSS, Associate Editor
University of Wisconsin-Madison

LEO POSTMAN, Associate Editor
University of California, Berkeley

DELOS D. WICKENS, Associate Editor
Ohio State University

Consulting Editors

NANCY S. ANDERSON, *University of Maryland*
NORMAN H. ANDERSON, *University of California, San Diego*
WILLIAM H. BATCHELDER, *University of California, Irvine*
WILLIAM F. BATTIG, *University of Colorado*
LYLE E. BOURNE, JR., *University of Colorado*
GEORGE E. BRIGGS, *Ohio State University*
E. J. CAPALDI, *Purdue University*
VIRGIL R. CARLSON, *National Institute of Mental Health*
HERBERT H. CLARK, *Stanford University*
ANDREW CRIDER, *Williams College*
HELEN B. DALY, *State University of New York, College at Oswego*
ROGER L. DOMINOWSKI, *University of Illinois at Chicago Circle*
SHELDON M. EBENHOLTZ, *University of Wisconsin-Madison*
WARD EDWARDS, *University of Michigan*
HOWARD EGETH, *Johns Hopkins University*
BRUCE R. EKSTRAND, *University of Colorado*
HENRY C. ELLIS, *University of New Mexico*
JAMES R. ERICKSON, *Ohio State University*
MURRAY GLANZER, *New York University*
JUDITH GOGGIN, *University of Texas at El Paso*
FRANCES K. GRAHAM, *University of Wisconsin*
WILLIAM W. GRINGS, *University of Southern California*
ROBERT C. HAYGOOD, *Arizona State University*
ELIOT HEARST, *Indiana University*
JULIAN HOCHBERG, *Columbia University*
LEONARD M. HOROWITZ, *Stanford University*
WILLIAM C. HOWELL, *Rice University*
STEWART H. HULSE, *Johns Hopkins University*
GEOFFREY KEPPEL, *University of California, Berkeley*
GREGORY A. KIMBLE, *University of Colorado*
HERBERT D. KIMMEL, *University of South Florida*
PATRICK R. LAUGHLIN, *University of Illinois*
HERSCHEL W. LEIBOWITZ, *Pennsylvania State University*
FRANK A. LOGAN, *University of New Mexico*
IRVING MALTZMAN, *University of California, Los Angeles*
DONALD R. MEYER, *Ohio State University*
BENNET B. MURDOCK, JR., *University of Toronto*
RAYMOND S. NICKERSON, *Bolt, Beranek, and Newman, Inc., Cambridge, Massachusetts*
MERRILL E. NOBLE, *Pennsylvania State University*
J. BRUCE OVERMIER, *University of Minnesota*
MICHAEL I. POSNER, *University of Oregon*
WILLIAM F. PROKASY, *University of Utah*
L. STARLING REID, *University of Virginia*
DONALD A. RILEY, *University of California, Berkeley*
WILLARD N. RUNQUIST, *University of Alberta*
RUDOLPH SCHULZ, *University of Iowa*
BRYAN E. SHEPP, *Brown University*
J. RICHARD SIMON, *University of Iowa*
PAUL SLOVIC, *Oregon Research Institute*
JANET T. SPENCE, *University of Texas at Austin*
HERBERT H. STENSON, *University of Illinois at Chicago Circle*
DAVID R. THOMAS, *University of Colorado*
MILTON A. TRAPOLD, *University of Minnesota*
ENDEL TULVING, *Yale University*
WILLIAM P. WALLACE, *University of Nevada*
WAYNE A. WICKELGREN, *University of Oregon*
GORDON WOOD, *Michigan State University*

HAROLD P. VAN COTT
Managing Editor

RITA JASKOWITZ
BETH FORRESTER
Technical Editors

HENRY S. SALMON
Advertising and Business Manager

ANNE REDMAN
Subscription Manager

Volume 95, 1972

Bureau ... Psy ...
S. C. E ...
6.3.81
Acc 5708

PUBLISHED MONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA. 17604
AND 1200 SEVENTEENTH ST. N. W., WASHINGTON, D. C. 20036

Second-class postage paid at Lancaster, Pa.

CONTENTS OF VOLUME 95

Agdern, M. See Girgus, J. S.

Allen, G. A. Memory Probes during Two-Choice Differential Reward Problems 78

Axelrod, S. See Guzy, L. T.

Bartlett, N. R. See Best, P. R., Jr.

Barton, A. K., and Young, R. K. Transfer from Free-Recall to Paired-Associate Learning .. 240

Baumeister, A. A. See Berry, F. M.

Begg, I., and Rowe, E. J. Continuous Judgments of Word Frequency and Familiarity 48

Bell, B. See Groninger, L. D.

Berry, F. M., Joubert, C. E., and Baumeister, A. A. Stimulus Selection and Meaningfulness at Different Stages of Paired-Associate Learning 188

Best, P. R., Jr., and Bartlett, N. R. Effect of Stimulus Interval and Foreperiod Duration on Temporal Synchronization 154

Borges, M. Increasing Item Accessibility in Free Recall 66

Bradshaw, J. L. See Geffen, G.

Brickman, P., and Pierce, S. M. Estimates of Conditional Probabilities of Confirming versus Disconfirming Events as a Function of Inference Situation and Prior Evidence .. 235

Brown, A. L., and Campione, J. C. Recognition Memory for Perceptually Similar Pictures in Preschool Children 55

Burnside, B. L. See Restle, F.

Campione, J. C. See Brown, A. L.

Carre, F. A. See Hintzman, D. L.

Cole, L. E., and Kanak, N. J. Transfer of Implicit Associative Responses between Free-Recall Learning and Verbal Discrimination Learning Tasks 110

Coren, S. See Girgus, J. S.

Cox, S., and Dember, W. N. U-Shaped Metacontrast Functions with a Detection Task 327

Champion, R. A., McCann, T. E., and Ruffels, J. A. Frustration Phenomena in Paired-Associate Learning 123

Chatfield, D. C., and Janek, E. J. Attribute Selection in Concept Identification 97

Clark, J. L., Warm, J. S., and Schumsky, D. A. General and Specific Factors in the Intersensory Transfer of Form 183

Curtis, D. W., and Rule, S. J. Magnitude Judgments of Brightness and Brightness Difference as a Function of Background Reflectance 214

Cymer, W. See Groninger, L. D.

del Castillo, D. M., and Gumenik, W. E. Sequential Memory for Familiar and Unfamiliar Forms 90

Dember, W. N. See Cox, S.

Deregowski, J., and Ellis, H. D. Effect of Stimulus Orientation upon Haptic Perception of the Horizontal–Vertical Illusion 14

Derks, P. L. Visual Recognition of Similarity and Identity 237

Detterman, D. K., and Ellis, N. R. Determinants of Induced Amnesia in Short-Term Memory 308

Dickerson, D. J., Novik, N., and Gould, S. A. Acquisition and Extinction Rates as Determinants of Age Changes in Discrimination Shift Behavior 116

Dieter, R. See Teller, K. J.

Dillon, R. F. See Petrusic, W. M.

Di Lollo, V. See Hogben, J.

Dolinsky, R. Clustering and Free Recall with Alternative Organizational Cues 159

Dominowski, R. L. Effects of Solution Familiarity and Number of Alternatives on Problem Difficulty 223

Ebert, T. H., and Fallon, D. Recall and Resistance to Unlearning of Verbal Mediating Associates as a Function of Anticipation Interval 251

Elias, M. F., and Kinsbourne, M. Time Course of Identity and Category Matching by Spatial Orientation 177

Ellis, H. D. See Deregowski, J.

Ellis, N. R. See Detterman, D. K.

Epstein, W., and Wilder, L. Searching for To-Be-Forgotten Material in a Directed Forgetting Task 349

Eskridge, V. L. See Hintzman, D. L.

Evans, A. See Runquist, W. N.

Falkenberg, P. R. Recall Improves in Short-Term Memory the More Recall Context Resembles Learning Context 39

Fallon, D. See Ebert, T. H.

Fattaleh, D. L. See Schwartz, S. H.
Ferris, S. H. Motion Parallax and Absolute Distance 258
Fink, R. T. Response Latency as a Function of Hypothesis-Testing Strategies in Concept Identification 337
Fozard, J. L., and Weinert, J. R. Absolute Judgments of Recency for Pictures and Nouns after Various Numbers of Intervening Items 472
Frost, N. Encoding and Retrieval in Visual Memory Tasks 317
Furedy, J. J., and Klajner, F. Preference for Information about an Unmodifiable but Rewarding Outcome 469
Geffen, G., Bradshaw, J. L., and Nettleton, N. C. Hemispheric Asymmetry: Verbal and Spatial Encoding of Visual Stimuli 25
Gianutsos, R. Free Recall of Grouped Words 419
Girgus, J. S., Coren, S., and Agdern, M. The Interrelationship between the Ebbinghaus and Delboeuf Illusions 453
Giurintano, S. L. Directionality of Associations in Paired-Associate Learning 463
Godek, C. L. See Greene, R. T.
Godkewitsch, M. The Role of Language Habits in Understanding Letter Sound Sequences 63
Gorfein, D. S., and Jacobson, D. E. Proactive Effects in Short-Term Recognition Memory 210
Gould, S. A. See Dickerson, D. J.
Grant, D. A. See Ornstein, P. A.
Groninger, L. D., Bell, B., Cymer, W., and Wess, B., Jr. Storage Aspects of Nouns Presented under Imagery and Acoustic Coding Instructions 194
Greene, R. T., Lawson, R. B., and Godek, C. L. The Ponzo Illusion in Stereoscopic Space .. 358
Gumenik, W. E. See del Castillo, D. M.
Guzy, L. T., and Axelrod, S. Interaural Attention Shifting as Response 290
Hintzman, D. L., Carre, F. A., Eskridge, V. L., Owens, A. M., Shaff, S. S., and Sparks, M. E. "Stroop" Effect: Input or Output Phenomenon? 458
Hogben, J., and Di Lollo, V. Effects of Duration of Masking Stimulus and Dark Interval on the Detection of a Test Disk 245
Howell, W. C. Compounding Uncertainty from Internal Sources 6
Humphreys, M. S., Petrusic, W. M., and Schwartz, R. M. Free Recall following a Switch in Encoding Class 455
Jacobson, D. E. See Gorfein, D. S.
Janek, E. J. See Chatfield, D. C.
Joubert, C. E. See Berry, F. M.
Kanak, N. J. See Cole, L. E.
Kantowitz, B. H. Interference in Short-Term Motor Memory: Interpolated Task Difficulty, Similarity, or Activity? 264
Kaplan, I. T., Metlay, W., and Lyons, C. T. Display Size and the Distribution of Search Times 334
Kaplan, K. J. Vicarious Reinforcement and Model's Behavior in Verbal Learning and Imitation 448
Kendler, H. H., Kendler, T. S., and Ward, J. W. An Ontogenetic Analysis of Optional Intradimensional and Extradimensional Shifts 102
Kendler, T. S. See Kendler, H. H.
Kinsbourne, M. See Elias, M. F.
Kirsner, K. Naming Latency Facilitation: An Analysis of the Encoding Component in Recognition Reaction Time 171
Klajner, F. See Furedy, J. J.
Kroll, N. E. A., Parkinson, S. R., and Parks, T. E. Sensory and Active Storage of Compound Visual and Auditory Stimuli 32
Krueger, L. E. Sentence–Picture Comparison: A Test of Additivity of Processing Time for Feature Matching and Negation Coding 275
Lawson, R. B. See Greene, R. T.
Levin, H. S. See Levis, D. J.
Levis, D. J., and Levin, H. S. Escape Maintenance under Serial and Simultaneous Compound Presentations of Separately Established Conditioned Stimuli 451
Lewis, P. See Poetter, R. A.
Ley, R. See Locascio, D.
Liu, I.-M. Effects of "Right" and "Wrong" in a Thorndikian Experiment 365
Locascio, D., and Ley, R. Associative Reaction Time, Meaningfulness, and Mode of Study in Free Recall 460
Loftness, S. See Sadalla, E. K.
Lyons, C. T. See Kaplan, I. T.
Marchant, H. G., III, Mis, F. W., and Moore, J. W. Conditioned Inhibition of the Rabbit's Nictitating Membrane Response 408

Marshall, S. H. Recognition and Recall in Short-Term Motor Memory 147
McCann, T. E. See Champion, R. A.
McCarthy, S. V. Visual Serial Search Performance for Number and Letter Targets 233
McCormack, P. D., and Swenson, A. L. Recognition Memory for Common and Rare Words 72
McNish, R. D. See Shontz, F. C.
Metlay, W. See Kaplan, I. T.
Meunier, G. F., Ritz, D., and Meunier, J. A. Rehearsal of Individual Items in Short-Term Memory 465
Meunier, J. A. See Meunier, G. F.
Mis, F. W. See Marchant, H. G., III
Moore, J. W. See Marchant, H. G., III
Myer, B. M., and O'Connell, D. C. Memory Span: Effects of String Length and String Composition 231
Nelson, T. O., and Smith, E. E. Acquisition and Forgetting of Heirarchically Organized Information in Long-Term Memory 388
Nettleton, N. C. See Geffen, G.
Nickerson, R. S. Auditory Codability and Short-Term Retention of Visual Information .. 429
Novik, N. See Dickerson, D. J.
O'Connell, D. C. See Myer, B. M.
Ornstein, P. A., Grant, D. A., and Watters, W. C. Semantic Generalization over a Bipolar Dimension of Meaning 201
Osborne, J. W. Interaction of Arousal and Number of Learning Trials in Paired-Associate Learning 135
Owens, A. M. See Hintzman, D. L.
Parkinson, S. R. See Kroll, N. E. A.
Parkman, J. M. Temporal Aspects of Simple Multiplication and Comparison 437
Parks, T. E. See Kroll, N. E. A.
Petrusic, W. M. See Humphreys, M. S.
Petrusic, W. M., and Dillon, R. F. Proactive Interference in Short-Term Recognition and Recall Memory 412
Pierce, S. M. See Brickman, P.
Poetter, R. A., and Lewis, P. Facilitation of Human Operant Responding by Stimuli which Precede Aversive Events 382
Restle, F., and Burnside, B. L. Tracking of Serial Patterns 299
Ritz, D. See Meunier, G. F.
Robbins, D., and Wise, P. S. Encoding Variability and Imagery: Evidence for a Spacing-Type Effect without Spacing 229
Robinson, G. M. The Delayed Auditory Feedback Effect Is a Function of Speech Rate 1
Rowe, E. J. Imagery and Frequency Processes in Verbal Discrimination Learning 149
Rowe, E. J. See Begg, I.
Ruffels, J. A. See Champion, R. A.
Rule, S. J. See Curtis, D. W.
Runquist, W. N., and Evans, A. Stimulus Recognition and Associative Coding 242
Sadalla, E. K., and Loftness, S. Emotional Images as Mediators in One-Trial Paired-Associate Learning 295
Sanford, A. J. Criterion Effects in Simple Reaction Time: Results with Stimulus Intensity and Duration Manipulations 370
Scarborough, D. L. Stmulus Modality Effects on Forgetting in Short-Term Memory 285
Schulz, L. S., and Straub, R. B. Effects of High-Priority Events on Recognition of Adjacent Items 467
Schumsky, D. A. See Clark, J. L.
Schwartz, R. M. See Humphreys, M. S.
Schwartz, S. H., and Fattaleh, D. L. Representative in Deductive Problem Solving: The Matrix 343
Shaff, S. S. See Hintzman, D. L.
Sheehan, P. W. Rule of Imagery in Incidental Learning: Replication and Extension of an Effect 226
Shontz, F. C., and McNish, R. D. The Human Body as Stimulus Object: Estimates of Distances between Body Landmarks 20
Sparks, M. E. See Hintzman, D. L.
Spitz, H. H. Effects of Redundancy Level and Presentation Method on the Paired-Associate Learning of Educable Retardates, Third Graders, and Eighth Graders 164
Smith, E. E. See Nelson, T. O.
Straub, R. B. See Schulz, L. S.
Suboski, M. D. See Teller, K. J.

Swenson, A. L. See McCormack, P. D.
Swensson, R. G. Trade-Off Bias and Efficiency Effects in Serial Choice Reactions 397
Teller, K. J., Dieter, R., and Suboski, M. D. Time Estimation and the Interstimulus Interval Function in Classical Conditioning 445
Uehling, B. S., and Underwood, B. J. Transfer as a Function of Stimulus, Response, and Simultaneous Stimulus and Response Similarity 375
Underwood, B. J. See Uehling, B. S.
Ward, J. W. See Kendler, H. H.
Warm, J. S. See Clark, J. L.
Watters, W. C. See Ornstein, P. A.
Weinert, J. R. See Fozard, J. L.
Wess, B., Jr. See Groninger, L. D.
Wilder, L. See Epstein, W.
Wise, P. S. See Robbins, D.
Young, R. K. See Barton, A. K.

Diary No 738(2)
Date 1-2
Lib
Bureau Edu Research.

Journal of
Experimental Psychology

VOL. 95, NO. 1 SEPTEMBER 1972

THE DELAYED AUDITORY FEEDBACK EFFECT IS A FUNCTION OF SPEECH RATE [1]

GEORGE M. ROBINSON [2]

Duke University

The delay producing maximal disruption of speech under delayed auditory feedback (DAF) is nearly equal to the duration of the average spoken syllable, supporting models of speech cognition with a fixed, internal time base of one-sixth of a second. By pacing *S*s with flashing lights, it was found possible to vary the speech rate under conditions of DAF. The delay maximally disrupting speech did not remain fixed but moved along as an inverse linear function of the speech rate, supporting a model of speech cognition that derives its timing from the speech itself and not from an internal clock. The results lessen the plausibility of the existence of a "psychological moment." Explanations for the DAF phenomenon are discussed.

Models for the perception, memory, and production of speech can be separated on the basis of whether they have an internal or an external source of timing. Internally timed models provide the speaker with a built-in clocking system independent of the durations of the elements of speech (Lenneberg, 1967). Externally timed models have no internal clock, but instead derive their timing from the (relative) durations of the speech elements themselves (Robinson, 1970). The first model is thus timed *independently* of the speech it is processing, while the timing of the second model depends directly upon the temporal pattern of the speech. The first would support and the second would deny a theory of discrete "psychological moments" as conceived by Stroud (1955) and White (1963).

Two experimental paradigms bear on this theoretical difference: the "alternation" experiment (Cherry, 1953) and the delayed auditory feedback (DAF) experiment (e.g., Black, 1951). In the first, *S* is required to "shadow" a recorded speech signal switched back and forth from one ear to the other. In the second, *S* is required to speak aloud while listening to his own voice fed back to him after being subjected to a short delay. Both the period of alternation in Cherry's experiment and the feedback delay in Black's experiment that led to *S*'s worst performance were about $\frac{1}{6}$ sec. This approximates the period of timing rhythm proposed by Lenneberg (1967, p. 110 ff).

However, Huggins (1964) demonstrated that when the speech played to *S* in the alternation experiments was speeded up, the period of alternation causing the worst performance was shorter. Therefore, the

---

[1] Thank you, David McNeill and Janice Moulton, for valuable criticism. National Science Foundation Grant GB 15369 supported the research.

[2] Requests for reprints should be sent to George M. Robinson, Department of Psychology, Duke University, Durham, North Carolina 27706.

© 1972 by the American Psychological Association, Inc.

maximally disruptive period of alternation depended upon the durations of the elements of the speech itself and was not an independent property of the speaker.

For the DAF paradigm, it would seem important to vary the speech rate to see whether the maximally disruptive delay would change along with it, analogous to the way Huggins (1964) varied the "listening" rate in the alternation paradigm. After several preliminary attempts, it soon became apparent why this had not been done before. It is extremely difficult for a person to maintain any constant speech rate while experiencing DAF. The strong "drive" to maintain coherent speech leads even the most disciplined *S* to gather his speech into rapid stacatto bursts and to perform other temporal distortions. This behavior itself suggests a relationship between speech rate and the maximally interfering delay, but it is difficult to measure. The following experiment overcame these difficulties and provided a strong test of the relation between syllable duration and delay length.

METHOD

Each of four *S*s sat alone inside an acoustic isolation chamber. Each wore earphones and faced an array of colored pilot lamps. A tape recording instructed *S* to count aloud into a microphone each flash of the lights. Each *S* was carefully instructed to enunciate each number clearly and evenly and to try to make the number last as long as the light stayed on. This technique was demonstrated for *S*s by the tape recording counting a sample series of light flashes.

The *S* then watched and counted five series of 50 flashes each, at each of eight different flash rates. The flash rate was constant within each series and increased monotonically across the eight series. In the order of presentation, the flash rates were as follows: one flash every 3.0, 2.0, 1.5, 1.0, .75, .50, and .20 sec. (a total of 2,000 flashes).

For the first 20 flashes of each series, *S* heard his own voice without delay through his earphones, accompanied by a background of "pink" noise (a white-noise spectrum restricted to the frequency band of human speech) of about 60 db SPL in order to mask that portion of his speech he could hear through the bones in his head.

Beginning with flash number 21, the feedback to *S* from his microphone to his earphones was delayed. The pink-noise background continued. The delay was achieved with a continuous-loop, variable-speed tape transport designed for this experiment.

The five different delays (see Fig. 2 abscissa) were distributed across the eight speech rates in an alternating ascending-descending series to provide 40 experimental conditions. The alternating series was used to prevent practice effects from favoring any particular combination of short or long delay with slow or fast speech rates.

The entire presentation to *S*, including directions, flashes, and switching between normal and delayed feedback was programmed on magnetic tape. The timing of the system including the delay unit was accurate to within 1%.

Scoring.—Each of the spoken numbers had a potential error score of 3, 1 point for each *whole* syllable missed ("thirty," "forty," and "fifty" only have two syllables, but they are balanced out by "twenty-seven," "thirty-seven," and "forty-seven," which have four). When a whole syllable was omitted or repeated, it scored 1 point, e.g., "thirtity-eight" and "for-five." When an error was made that affected only *part* of a syllable, it scored ½ point, e.g., inappropriate glottal stops like "twenty-si'ix," uncompleted syllables like "twenty-si," and striking quavers within a vowel like "forty-fiyiyive." An alternative scoring system, in which all errors were weighted the same, turned out only to *increase* the significance of the results to be discussed.

RESULTS

To permit comparison across speech rates, a normalized error score matrix was prepared for each *S* by setting the sum of the errors made at all five delays for each speech rate equal to 100 and proportioning the cells accordingly. The matrices appear in Table 1 along with the normalization factors by which the error scores were multiplied. The matrix for each *S* displays a strong diagonal tendency.

Table 2 combines the data for the four *S*s. An $8 \times 5 \times 4$ analysis of variance found a significant Speech Rate × Delay interaction, $F\ (28,\ 84) = 2.28,\ p < .01$. Delay alone was not a significant factor, $F\ (4,\ 12) = .51$.

Figure 1 displays the combined data, plotted to show articulation score as a function of feedback delay for three of the eight speech rates. The series of graphs clearly portray the disruptive effects of delayed auditory feedback as a function of speech rate.

While the maxima of the curves in Fig. 1 show a drift from right to left as the syllable duration decreases (speech rate increases), it would be valuable to know

TABLE 1

Normalized Articulation Error Scores with Normalization Factors for Four Subjects under Delayed Auditory Feedback

Rate (in sec/syllable)	Delay (in sec.)					Normalization factor	Delay (in sec.)					Normalization factor
	.360	.270	.180	.135	.090		.360	.270	.180	.135	.090	
	S_1						S_3					
1.0	40	24	16	16	4	4.0	57	29	14	0	0	14.3
.67	38	32	15	9	6	2.9	55	30	18	0	0	9.0
.50	44	38	13	6	0	6.3	19	38	25	6	13	6.25
.33	9	30	30	22	9	4.4	9	18	36	27	9	9.1
.25	67	10	6	12	4	2.1	38	0	13	25	25	12.5
.17	28	18	17	24	14	.63	13	21	25	25	16	.82
.10	14	14	15	26	32	.46	15	20	22	21	21	.45
.067	13	15	18	25	28	.33	18	20	17	23	21	.33
	S_2						S_4					
1.0	21	42	21	16	0	5.3	76	18	6	0	0	2.9
.67	31	31	13	13	13	6.3	51	36	12	0	0	2.4
.50	7	14	36	36	7	7.1	12	20	36	32	0	2.0
.33	7	36	21	21	14	7.1	16	25	35	23	1	1.2
.25	9	16	41	28	6	3.1	13	12	27	37	12	.41
.17	6	12	27	22	14	2.0	6	19	28	19	36	.78
.10	21	11	7	28	33	1.1	3	10	13	33	41	.52
.067	15	14	14	26	31	.53	3	10	13	31	43	.73

something of the function that relates the effects of DAF disruption to the interaction of syllable length and delay. Figure 2 shows the first-moment central tendency of articulation scores as a function of the delay for each speech rate. Admittedly, this is a crude and possibly misleading procedure for at least the two curves, which are truncated and do not with certainty return to the abscissa (i.e., the 1-sec. syllable rate and that for .67 sec/syllable, which is similar but displaced to the left). A least-squares line is cautiously offered. The function relating speech rate to the delay that is maximally disruptive to that rate is:

$$\text{Delay} = \tfrac{1}{6}\ \text{Mean Syllable Duration} + 160\ \text{msec.}$$

Note that the slope is an undimensioned number. It has no relationship to Lenneberg's (1967) "magical one-sixth second"

TABLE 2

Mean Normalized Error Score for Four Subjects as a Function of Syllable Length and Feedback Delay

Speech rate (in sec/syllable)	Delay (in sec.)										First-moment mean delay
	.360		.270		.180		.135		.90		
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	
1.0	48.5	20.4	28.3	8.8	14.3	5.4	8.0	8.0	1.0	1.7	290
.67	43.8	9.7	32.3	2.3	14.5	2.3	5.5	5.7	4.8	5.4	285
.50	20.5	14.2	27.5	10.7	27.5	9.5	20.0	14.1	5.0	5.4	240
.33	10.3	3.4	27.3	6.6	30.5	5.9	23.3	2.3	8.3	4.7	205
.25	31.8	23.2	9.5	5.9	24.3	15.0	23.0	6.4	11.8	8.2	215
.17	13.3	9.0	17.5	3.4	24.3	4.3	22.5	2.3	20.0	9.3	185
.10	13.3	6.5	13.8	3.9	14.3	5.4	27.0	4.3	31.8	7.2	175
.067	12.3	5.6	14.8	3.6	15.5	2.1	26.3	3.0	30.8	7.9	175

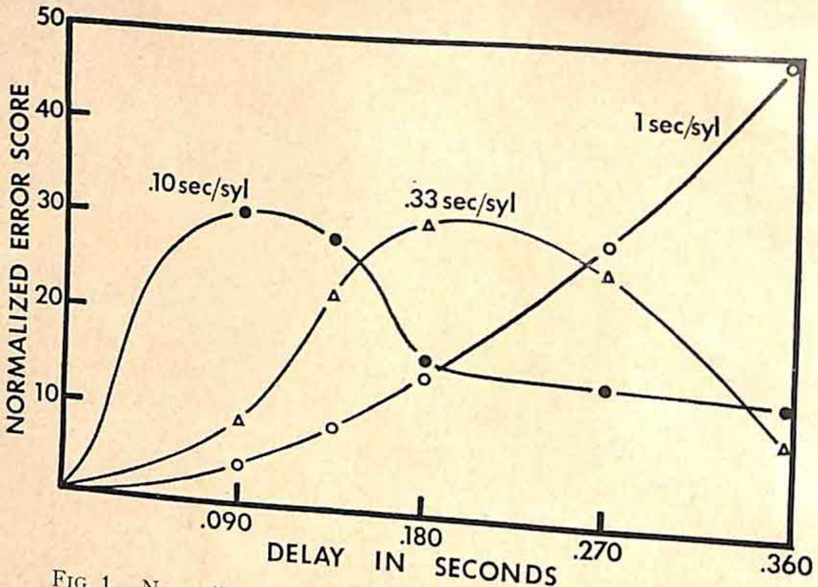


FIG. 1. Normalized articulation error score as a function of delay duration for three (typical) of the eight speech rates.

other than numerical coincidence. An alternative approach is to use a fractional-moment measure of central tendency and apply it only to those curves that promise to return to the abscissa. (As the fraction of the moment approaches zero, the measure of central tendency approaches the curve's maximum.) This procedure converges to a slope of around three.

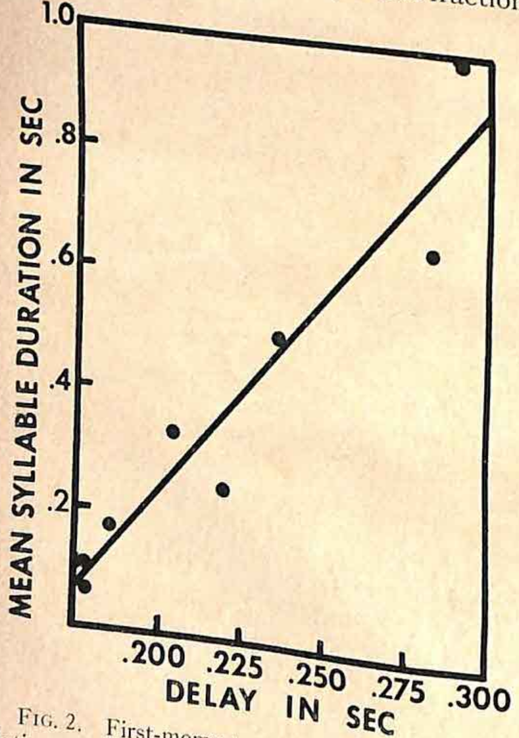


FIG. 2. First-moment central tendency of articulation score as a function of delay for each of eight speech rates.

DISCUSSION

Although the slope (and perhaps the shape of the function) cannot be precisely defined, the data unambiguously demonstrate that the delay that leads to maximal disruption of speech is a function of the speech rate. If the effect of DAF is related to the source of the timing of speech (both Lenneberg (1967) and the present author agree that it is), the results of this experiment favor a model whose timing depends on the speech itself, against one whose timing, built into the speaker, is independent of the speech.

Huggins (1964) reported that the median duration of spoken syllables was 180 msec. This is close to the feedback delay that maximally disrupts speech under "normal conditions." Such delay duration allows the greatest coincidence between the syllable being articulated and the auditory feedback from the previous syllable. The optimal overlap between spoken and heard syllables might confuse the speaker in one (or a combination) of two ways: first, the confusion arises from the discrepancy between the phonetic quality of what *S* is actually generating and what the feedback tells him he is producing. The *S* can avoid the discrepancy and regain much of the normal correspondence between spoken and heard syllables in two ways: he can repeat a syllable (stutter) or he can lengthen it (drawl).

Lee (1950), Fairbanks (1954), and Smith (1962) explained the DAF effect in the above way. However, within this general description, their models differ considerably. Lee stressed a hierarchy of servomechanical control loops; Fairbanks considered the several sources of feedback—proprioceptive, air-conducted, and bone-conducted sound; Smith incorporated the phenomenon into a "neurogeometric" theory (Smith & Smith, 1962), which stresses the interrelation of sensory and motor information at all levels instead of separating the motor and sensory aspects and permitting only one (or a few) discrete lines of negative feedback.

A second way in which DAF may disrupt speech is by misinforming *S* as to the temporal position of any particular syllable in the utterance. This is consistent with the "recirculation" theory suggested by Chase (1958) and with the "control mechanism" proposed by Huggins (1968). Whenever *S* relies on feedback from his articulatory apparatus, auditory feedback informs him that he is speaking too slowly. Speeding up, however, increases the (relative) temporal discrepancy between motor and sensory feedback. Abandoning this strategy, *S* then tries to pace himself by relying on the auditory feedback. But now motor feedback informs him that he has "gotten ahead of himself." Slowing down reduces the relative temporal discrepancy and leads to drawling, the prominent symptom of DAF. Escape from DAF does result when *S* slows down enough so that the duration of each syllable is large compared with the delay period. However, the demands of the experiment force him to speed up; most experiments required *S*s to maintain "normal" speech. The present experiment, by providing external pacing, was very rigorous in this demand. When *S* speeds up he finds himself again increasing the discrepancy between spoken and heard material and the cycle continues. We might think of *S* as trapped by DAF in a hysteresis loop of articulation rate as a function of the temporal discrepancy between sensory and motor feedback.

Earlier, Huggins (1968) tested the hypothesis now supported by the present experiment by requiring one *S* to read a sentence composed of all short syllables and one of all long syllables under different delays of auditory feedback. Huggins failed to find a significant interaction between delay length and syllable duration probably because he permitted his *S* to change the duration of syllables by as much as 250% and because he did not mask bone-conducted feedback. However, his "fine analysis" of the spectrograms of his *S*'s speech led him to suggest that DAF disrupts speech by interfering with the ability of *S*'s control mechanism to "...keep track of its 'place' in the program [of motor instructions for speech] [p. 59]." This is more consistent with the idea that the DAF effect results from misinformation as to the temporal position rather than as to the phonetic quality of the syllable.

REFERENCES

Black, J. W. The effects of delayed side-tone upon vocal rate and intensity. *Journal of Speech and Hearing Disorders*, 1951, **16**, 56–60.

Chase, R. A. Effect of delayed auditory feedback on the repetition of speech sounds. *Journal of Speech and Hearing Disorders*, 1958, **23**, 583–590.

Cherry, E. C. Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 1953, **25**, 975–979.

Fairbanks, G. Systematic research in experimental phonetics: I. A theory of the speech mechanism as a servosystem. *Journal of Speech and Hearing Disorders*, 1954, **19**, 133–139.

Huggins, A. W. F. Distortion of the temporal pattern of speech: Interruption and alternation. *Journal of the Acoustical Society of America*, 1964, **36**, 1174–1178.

Huggins, A. W. F. Delayed auditory feedback and the temporal properties of speech material. *Zeitschrift fur Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 1968, **21**, 54–60.

Lee, B. S. Effects of delayed speech feedback. *Journal of the Acoustical Society of America*, 1950, **22**, 824–826.

Lenneberg, E. *Biological foundations of language*. New York: Wiley, 1967.

Robinson, G. M. *Rhythm in verbal memory: A theoretical model and an experimental basis:* Unpublished doctoral dissertation, University of Chicago, 1970.

Smith, K. U. *Delayed sensory feedback and behavior*. Philadelphia: Saunders, 1962.

Smith, K. U., & Smith, W. M. *Perception and motion: An analysis of space structured behavior*. Philadelphia: Saunders, 1962.

Stroud, J. The fine structure of psychological time. In H. Quastler (Ed.), *Information theory in psychology*. New York: Free Press, 1955.

White, C. T. Temporal numerosity and the psychological unit of duration. *Psychological Monographs*, 1963, **77**, (2, Whole No. 575).

(Received February 1, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 6–13

COMPOUNDING UNCERTAINTY FROM INTERNAL SOURCES[1]

WILLIAM C. HOWELL[2]

Rice University

Subjective probability (SP) for skill-based or *internal* events has been shown to differ in certain ways from the more conventional SP for environment-based or *external* events. The present research extended two other issues from the external SP literature to the internal case: (*a*) comparison of direct versus choice-inferred measures of SP and (*b*) description of rules for intuitive combination of elemental SP values. Additionally, range of uncertainties was extended to include much lower values than before, and a recent study of preferences among internal–external compound events was replicated for triple compounds. Findings generally supported earlier work in the consistent tendency toward overconfidence. As in the external case, direct and indirect SP measures were highly correlated, although the former were lower and apparently more valid. People seemed to combine SP values intuitively in a predominantly multiplicative fashion.

When people are required to deal with uncertainty arising from their own behavior (*internal* or skill-derived uncertainty), they show a consistent tendency toward overconfidence. This tendency has now been demonstrated under a number of experimental conditions for several very different kinds of behavior: a perceptual discrimination task (Howell, 1967), a probability learning situation (Howell, 1970), and a choice task involving combinations of dart throwing and roulette (Howell, 1971). Only in the first of these studies was there an indication that *S* ever underestimates objective probability (OP) for internal events, and even this was restricted to the higher OP levels, where a ceiling limitation begins to enter the picture. Moreover, one could argue that from *S*'s standpoint the uncertainty in that task was not entirely self-generated—*S* could occasionally "do well" and still "lose." Thus the predominant tendency toward overestimation seems well established.

The significance of this generalization, as well as several other findings of more limited scope, lies in the fact that (*a*) they are not entirely consistent with typical *external* (environment-derived) uncertainty data, and (*b*) most of the empirical and theoretical work on decision making, choice behavior, subjective probability and the like is based exclusively on such external uncertainty (see, for example, reviews by Peterson & Beach, 1967; Slovic & Lichtenstein, 1971). It is doubtful, in view of discrepancies observed so far, that laws for external uncertainty can be applied without some adjustment to internal uncertainty, a component of which exists in nearly all extralaboratory decision situations.

The present research deals with two other areas in which source of uncertainty might be worthy of consideration: rules for combining the uncertainty in multiple or compound events, and comparability of subjective probability (SP) estimates derived from judgment versus choice.

Both of these areas have received considerable attention in the case of external sources. With a few notable exceptions, it has usually been found that people are conservative in combining probabilities intuitively; expressed in Bayesian terms, they fail to revise prior values in light of new evidence as much as they should (Slovic & Lichtenstein, 1971). On the other hand, SP seems a viable construct in that there is a high correlation between measures people produce directly through estimation and those inferred indirectly from their choice behavior (Beach

---

[1] This research was supported by Grant GB 8052 from the National Science Foundation. The assistance of Katherine E. DePitts, Deborah L. Vandell, Martha M. Van Parys, and Mary A. Falk in collecting and analyzing the data is gratefully acknowledged.

[2] Requests for reprints should be sent to William C. Howell, Department of Psychology, Rice University, Houston, Texas 77001.

& Phillips, 1967; Beach & Wise, 1969a). The only difference is that direct estimation seems to yield the more reliable and efficient measure (Beach & Wise, 1969b; Wise, 1970).

Neither of these conclusions has been extended to the case of internal uncertainty. An incidental finding from an earlier study suggested that the correlation between inferred and directly estimated SP may not be as high for internal as for external sources (Howell, 1970). If true, it would throw into question the practice of considering SP a unitary psychological dimension in choice and decision models.

Three studies were designed to address various aspects of the combination-rule and direct versus indirect measurement issues. In the first, both scaling techniques were applied to a set of internal events that included compound as well as simple occurrences. The SP values yielded by the two methods were compared, and compound event judgments were compared with estimates derived from simple judgments. Thus both issues were studied for events considered *individually*. The second and third experiments examined the consistency of preferences involving *pairs* of these events pitted against each other in a decision situation. The second dealt exclusively with internal uncertainty and the third, in an extension of an earlier experiment (Howell, 1971), combined internal with external sources in the definition of each (compound) event.

GENERAL METHOD

Subjects.—Thirty undergraduate volunteers, 17 males and 13 females, were paid $1.25/hr plus bonuses based on performance for participation in this research. Two of these, 1 male and 1 female, were dropped for failure to achieve any semblance of skill in dart throwing. All of the remaining 28 served successively in all three studies after having reached an asymptotic skill level on the "internal event" task. This procedure was dictated by the cumulative nature of the research; i.e., information gathered on each *S* and knowledge acquired by each *S* in the first experiment was necessary for the second and third.

General task.—The events for which internal uncertainty was measured (Exp. I) and upon which decisions were based (Exp. II and III) were dart throws for specified target criteria (rings on a dart board). A *simple* event was a single throw; a *compound* event was a sequence of two, three, or four throws. External uncertainty was provided by roulette-type spinners, again in simple (one spin) or compound (more than one spin) form. Both techniques are described in detail elsewhere (Howell, 1967, 1970, 1971). In the first two studies, both simple and compound events were comprised of dart throwing (internal uncertainty) exclusively, whereas mixtures of the two tasks (e.g., an event defined in terms of one throw and two spins) were included in the third study. In all three studies, *S* was required to make judgments or choices on the basis of how likely he considered a favorable outcome to be for each event as a unit: i.e., a single "hit" with a dart or spin in the case of simple events, a "hit" on all darts and/or spins prescribed for a compound event.

Pretraining.—Initial training in dart throwing was carried out in daily sessions of approximately 40 min. each until accuracy had stabilized on all criterion rings. The *S* threw four darts in succession at a 30.5-cm. rifle target mounted on fiberboard at a distance of about 2.9 m. He received feedback on total "score" for the four throws, although accuracy was calculated separately for each target ring over each session in order to estimate his asymptotic level. The number of sessions required varied from two to four.

Following dart training, two more sessions were devoted to decision training in which *S* was familiarized with a procedure for later use in scaling SP. This procedure involved a series of choices between a skill task (dart throwing) and a chance task (roulette spinning) for a fixed payoff of 1¢ on each favorable outcome. Each prescribed dart task was pitted against a series of spinner values ranging from .1 to .9, and the subjective indifference point, expressed in "objective" spinner values, was taken as an index of SP. A more detailed account of this procedure and its rationale are developed elsewhere (Howell, 1967, 1971). The internal events for which SP was measured by this means were single dart throws spanning the range of difficulty for each *S*. Accuracy was monitored here, too, in order to detect any marked shifts in asymptotic level.

EXPERIMENT I

The immediate purpose of this study was to establish SP values for a set of events, both simple and compound, all involving the dart-throwing task. How confident, it was asked, is *S* in his ability to throw a prescribed number of darts within prescribed criterion rings? The ultimate purpose was twofold: (*a*) to determine whether SP measured in the fashion just described (i.e., through choices involving an external "standard" scale) is comparable to that measured by direct rating, and (*b*) to explore the manner in which S combines simple SP values, however measured, into compound ones.

Design and Procedure

A pool of 14 events was constructed of all combinations of a difficult (H) and an easy (E) dart-throwing criterion taken singly (H, E), in double compounds (HH, EH, EE), triple compounds (HHH, HHE,...EEE), and quadruple compounds (HHHH,...EEEE). The criterion rings for H and E were selected for each individual so as to approximate as closely as possible OP (asymptotic accuracy) values of .50 and .80. Difficulty level for each event was thus equated across *S*s. Furthermore, a rough estimate of the comparative difficulty of the 14 events was possible under the assumption that each throw in a compound event is independent. On this basis, expected accuracy varied from .80 for the E event to about .07 for the HHHH event. *Actual* difficulty (without the independence assumption) was determined later for each S using accuracy (OP) data collected during the experiment proper. For compound events it was required that S throw for criteria in the order specified (e.g., in HEE, he had to hit the difficult ring on his first throw, the easy one on his last two). When compound events included both H and E components, internal order was dealt with as an extraneous variable and controlled within each sequence in which the event appeared; e.g., EH and HE appeared equally often in any sequence.

Ten of the 14 events were first scaled for SP indirectly as in pretraining, with each one being pitted against a series of nine spinner values four times (two ascending and two descending series).[3] The resulting 40 series (10 Events × 4 Series Replications) were presented in an order randomized separately for each S, and the mean indifference point was calculated for each S-event combination. Following this indirect estimation, SP was scaled again using a direct procedure. Here all 14 events were presented twice each, in a random order, and S estimated the accuracy of his performance on each one in percent correct. Again, means for S-event combinations were computed. The reason for scheduling the direct after the indirect scaling in all cases was to reduce the possibility of a spurious correlation that would result if S attempted to adjust his choice behavior in accordance with prior numerical estimates. The reverse bias, adjusting numerical estimates according to prior choice behavior, seemed a far less serious danger in that S would scarcely be expected to remember—when numbers were called for—how he had responded on the 360 preceding decisions. He might, of course, have acquired some knowledge about relative frequencies; but that would be entirely consistent with the objectives of the research and thus hardly a "bias."

Results

For simple and compound events alike, whether directly or indirectly measured, the tendency again was solidly in the direction of overconfidence. This is seen most clearly in Fig. 1, a plot of the SP–OP difference scores by the two scaling techniques for all 10 conditions in which both methods were used. The OP values in all cases were the empirical accuracy scores obtained over the 36 occasions on which S performed each event (single or compound). Here, as in all cases where average data are presented, computations were carried out for individual *S*s and then combined. The trend, as in earlier work, was for overconfidence to vary inversely with OP—especially for the choice-derived measure. Variability of the points in Fig. 1, incidentally, was large but comparable for both functions: the semi-interquartile range was about .06 for the indirect and .07 for direct measures. The distribution of indirect measures was markedly skewed.

Figure 1 also shows that direct scaling yielded consistently lower SP estimates than did the indirect method; median overconfidence measured directly ranged from about .05 to .19, while the comparable range for the choice-derived measure was about .09 to .28. Correlation between the two sets of data, however, was high ($r = .83$).

The above conclusions were substantiated by an analysis of variance applied to median scores for each S. Both the difference between measures, F (1, 540) = 48.27, and the overall trend over OP, F (9, 540) = 7.89, were significant at $p < .01$.

Turning to the question of how people *combine* internal uncertainty information, the relevant data are summarized in Table 1. In the first two columns, mean OP (accuracy) and SP are presented for all 14 kinds of events. Data collected by the indirect method did not cover the full range of events and hence were omitted here; they were, however, included in all analyses where selected subsets were of

---

[3] The events eliminated were those least suitable for use in the subsequent studies.

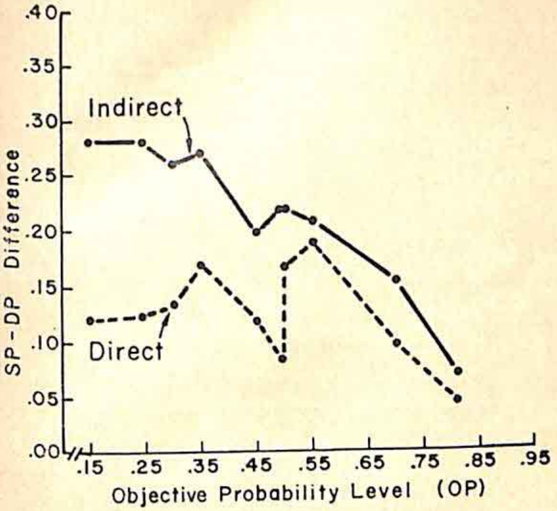


FIG. 1. Comparison of the overconfidence (SP–OP) functions for the two ways of measuring SP.

interest. The last two columns give theoretical values for the same events based upon the assumption that *S* deals with the component throws as though they were independent. These predicted values were developed on the basis of each individual's performance on the simple events (H and E). In other words, predicted OP for a compound (say EH) was computed by multiplying the accuracy scores made by each *S* on the simple events of which it was composed (e.g., .81 × .52) and averaging them. Predicted SP was derived in the same fashion using *S*'s own judgments on E and H events as the basis for computation. The rationale for such an approach is that measured SP for a simple event is our best estimate of *S*'s confidence in a single throw. If we compare measured SP for a compound event (his demonstrated confidence in a prescribed multiple throw) with the product of his single-throw confidence values, we can deduce how much of his confidence in success on the compound event is contributed by factors associated with compounding per se. Thus, in a sense, we can factor out the overconfidence that we know accompanies the component acts and examine what is left. If there were no residual, i.e., if obtained scores agreed with predicted ones, it would suggest that *S* considers the uncertainty in elements of a compound to be independent and aggregates accordingly. Should the residual change systematically as more throws are added to the compound, it would suggest that he does not consider the uncertainties independent. Other possibilities, of course, also exist.

One such possibility is what actually occurred. Looking at Table 1, it is clear that there was a residual SP attributable to some facet of compounding: obtained SP was consistently above that predicted by the straight independence model. On an individual basis, 72% of the obtained values exceeded those predicted, $\chi^2 = 64.80$, $p < .001$ on the null hypothesis of 50%, as did 65% of the choice-derived values, $\chi^2 = 18.64$, $p < .001$. Thus the net effect of compounding seems to have been a further increased level of overconfidence. However, the data in Table 1 failed to show any obvious trend in this residual overconfidence, and several further analyses were undertaken.

Analyses of variance were performed to determine whether the magnitude of the combination effect did, in fact, vary as a function of (*a*) number of throws combined, (*b*) H–E composition of these compounds, and (*c*) OP level. Since OP is necessarily confounded with the other two variables, it was decided to carry out the

TABLE 1

SUMMARY OF OBTAINED AND PREDICTED ACCURACY (OP) AND CONFIDENCE (SP) LEVELS FOR ALL EVENTS

Event	Obtained		Predicted	
	OP	SP	OP	SP
E	.81	.86	—	—
EE	.70	.81	.66	.75
EEE	.55	.74	.53	.65
EEEE	.50	.67	.43	.56
H	.52	.60	—	—
HE	.45	.57	.42	.52
HEE	.35	.52	.34	.45
HEEE	.28	.48	.28	.39
HH	.31	.45	.27	.36
HHE	.25	.38	.22	.31
HHEE	.18	.39	.18	.27
HHH	.15	.27	.14	.21
HHHE	.11	.24	.11	.19
HHHH	.07	.18	.07	.13

Note.—Prediction based upon the product of probability values measured for component simple events.

analysis in two stages on selected conditions. In the first, number (2, 3, or 4 components) was crossed with composition (all-H, mixed, all-E) and OP varied freely. In the second, number was crossed with OP using those compound events—however constituted—which gave comparable OP across number. It was possible by this means to cross the three number levels with two OP levels (high and low). Results of these analyses were uniformly nonsignificant ($p > .05$) with one exception: the Number × Composition interaction for the direct estimation index, $F(4, 243) = 5.22$, $p < .01$. Table 2 shows this effect to be a differential number trend for the all-H versus the other composition conditions. For all-H compounds, the combination effect decreased with number while the all-E and mixed compounds it increased.

While statistically significant, this interaction effect was not particularly large in an absolute sense. Neither is it easily interpreted, for both the level and range of OP varied concomitantly: from roughly .07 to .30 for all-H; .50 to .70 for all-E. Thus, while a reliable but small source of systematic variance probably does exist, it seems far more realistic to emphasize the remarkable stability of the combination effect considering the wide variety of specific compounds studied. It seems as though *S* adds to his SP a fairly constant adjustment for compounds, whatever their nature, rather than some differential bias to successive components or steps in the aggregation process. Were the latter the case, the effect would have been a consistent trend over both number of components and OP. This contention is further supported by an analysis of the average "impact" of each event on all higher order events in which it appears. The impact of an event in any higher order event (e.g., of HE in HEE) is calculated by considering the higher order event a product of two components, the lower order one (HE) and a multiplier (E); then SP measured for the higher order event (HEE) is simply divided by SP for the multiplier (E). This, in effect, indicates how much impact SP_{HE} must have had on SP_E to produce SP_{HEE} assuming the multiplicative rule. Average impact estimates should approximate obtained SP values to the extent that *S*s tend to combine uncertainties in a multiplicative fashion. Comparison of these two measures is given in Table 3. Considering that the average impact estimates were not adjusted for the postulated combination effect, and higher order events were necessarily based on fewer cases than lower order ones, the agreement seems quite good.

TABLE 2

Mean Overestimation of SP Attributable to the Combination Process (Selected Conditions; Direct Measures)

Composition	No. combined			
	2	3	4	$\bar{X}$
All-H	.11	.10	.05	.09
Mixed	.04	.06	.12	.07
All-E	.06	.09	.10	.08
$\bar{X}$	.07	.08	.09	.08

TABLE 3

Comparison of Estimated "Impact" of Components on Higher Order Events with Directly Measured SP for These Same Components

Component	Mean impact	Direct SP
H	.66	.60
E	.91	.86
HE	.60	.57
EE	.85	.81
HH	.44	.45
HEE	.62	.52
HHE	.42	.38
HHH	.29	.27
EEE	.79	.74

EXPERIMENT II

The purpose of this experiment was to study the validity of the uncertainty estimates obtained in Exp. I by using the same events in a choice paradigm. Specifically, *S* was required to choose between pairs of events—all defined as simple or compound dart throws—for a fixed payoff. Since all uncertainty was internal, he would be expected to show a preference for the event with the higher measured SP (if these measures are valid). Moreover, the predictive accuracy of the two SP measures should differ if one is more valid than the other.

Design and Procedure

Fifteen pairs of events were presented to each of the 28 *S*s twice in a random order. The pairs, which were drawn from the set of 14 events described in Exp. I, listed in approximate order of decreasing SP difference were as follows:

Event A	Event B
H	EEE
EEEE	HE
HH	HEEE
HHEE	HHH
HEH	HEEE
HEE	HH
EEEE	H
HEE	EEEE
EEE	HE
HHH	HEEE
EE	H
EEEE	HH
HEH	EEEE
HH	EEE
EEEE	HHH

This subset of possible pairs was all that remained after eliminating those *S* could handle easily without considering SP: i.e., all pairs in which one member (EE) was a component of the other (EEE, HEE, etc.) or in which both had the same number of elements (hence one had more H's than the other—e.g., HHEE vs. HHHE). Since H is obviously harder than E, and it is obviously harder to hit $N + 1$ than N throws, choices involving such pairs would prove nothing.

All choice alternatives were presented verbally in a single session; on half of the occasions one member of each pair was read first, and on the other half it was read last. After each pair of alternatives was read out, *S* made his decision, also verbally, and threw the prescribed number of darts. His choice and success on that event were recorded; feedback, of course, was immediate. Following the last choice his total number correct was converted to cents, and he was paid off.

Results

The primary index of performance was the proportion of times the event with the larger subjective probability of success was preferred. For comparative purposes, both measures of SP were used. The value of each event on both of these scales was computed separately for each *S*. Differences between alternatives on each pair were then ordered and grouped into ranges, each range containing an approximately equal number of cases. This resulted in 9 different categories for indirectly measured SP and 10 for directly measured SP.

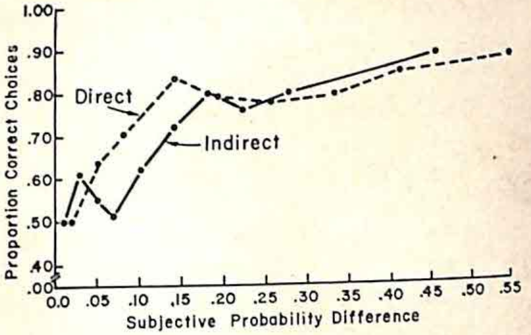


FIG. 2. Proportion of times the event with the higher SP was preferred at various SP difference levels. (Abscissa points are located at the midpoint of each SP difference interval.)

The two functions thus obtained are shown in Fig. 2. The *S*s appear to have been highly consistent in choosing the subjectively more likely event. The high correlation between SP measures obtained in Exp. I is reflected in the similarity of Fig. 2 functions. The only notable discrepancy occurs for differences below .20 where the direct measure indicates slightly better discrimination. In a more rigorous test of the superiority of the direct measure, a separate analysis was carried out on those decisions for which the indirect and direct measures led to *different* predictions. While only 107 of the 840 predictions were discrepant, (another reflection of the high correlation), on 60% of these occasions the direct measure was correct in predicting the decision. The resulting $\chi^2 = 4.56$, $p < .05$, is rather strong evidence of the superiority of direct over indirect SP methods, especially when one considers that some of the 107 cases were undoubtedly in the range of very small "real" SP differences where near-chance predictions would be expected.

EXPERIMENT III

Elsewhere it has been shown that in a choice between equiprobable events composed of an internal and external uncertainty component, *S* consistently prefers the alternative with the preponderance of uncertainty in the internal component (Howell, 1971). This follows from the generalization that overconfidence is more typical of SP for internal than for external events.

In studying that problem, it was necessary to deal with compound events, all of which had one internal and one external component. The present study was simply a replication of that study for triple compounds at four levels of combined OP: HHH, HHE, HEE, and EEE. In each case, 0, 1, 2, or 3 of the components were replaced by an external (roulette) component to provide an internal–external continuum. The entire range of difficulty was lower (about .12 to .51) than in the previously reported study (about .30 to .70).

Design and Procedure

Four levels of event difficulty (OP) were combined factorially with four levels of internal–external composition to yield 16 distinct triple compounds. In all cases, positioning of the internal and external components within each event was randomized.

TABLE 4

Choice Frequencies for Both Alternatives in Event Pairs Differing Only in Internal–External Composition

I-E			Difficulty level			
Pair	Alternative	Composition	HHH	HHE	HEE	EEE
1	a	100 I– 0 E	32	35	33	41
	b	0 I–100 E	24	21	23	15
2	a	67 I– 33 E	22	31	36	43
	b	33 I– 67 E	34	25	20	13
3	a	67 I– 33 E	38	22	31	40
	b	0 I–100 E	18	34	25	16
4	a	67 I– 33 E	25	30	29	39
	b	33 I– 67 E	31	26	27	17
5	a	100 I– 0 E	26	37	42	42
	b	67 I– 33 E	30	19	14	14
6	a	33 I– 67 E	35	34	33	41
	b	0 I–100 E	21	22	23	15

Note.—I-E = Internal–External. Compositions are given in percentages. Each I–E pair was repeated at four difficulty levels. Thus for Pair 1, Alternative a was preferred to Alternative b 32 times at difficulty HHH; b was preferred to a 15 times at difficulty EEE; etc.

Pairs of event alternatives were then formed from all combinations of events at each level of OP. Thus, for example, there were six HHH pairs, all differing in terms of the internal–external composition of the compounds.

The *S*'s task was to choose, in exactly the same fashion as Exp. II, which of each pair he preferred to "play" for a 1¢ payoff. The entire set of 24 pairs was presented twice to each *S* in a random order during a single session.

Results

The main interest was in how often choices favored the hypothesis that *S* prefers internal to external uncertainty sources. In other words, how often did he prefer the predominantly internal compound event? The answer to this question is given in Table 5. Over all *S*s and conditions, 61% of the choices favored the internal preference hypothesis. While this is significantly greater than would be expected by chance, $\chi^2 = 62.58$, $p < .001$, it is well below the 87% reported earlier.

One source of the discrepancy obviously lies in the difficulty variable. Whereas this proved to be a nonsignificant factor over the range of .30 to .70 covered by the earlier study, it was apparently quite important over the lower range studied here: only 53% of the decisions were in the hypothesized direction for the most difficult level (HHH; OP = .12), compared to 73% for the easiest level (EEE; OP = .51).

What accounts for the rest of the discrepancy is not entirely clear, although it could be simply that triple compounds are less easily discriminated on the basis of SP than are double compounds of the sort used previously.

General Discussion

Several conclusions seem justified in view of the foregoing results. First, evidence continues to mount regarding man's consistent overconfidence in uncertain events associated wholly or in part with his own performance. All three of the studies added further credence to this generalization by extending the range of conditions over which it holds to compound events, very low probability events, and direct measurement procedures. It is still debatable whether measured overconfidence represents subjective misperception of uncertainty per se.

or the operation of some other process—perhaps even a utility factor—which, together with SP and other utility factors, determine decision behavior. The fact that direct estimation of SP shows this tendency (and not just the more suspect choice-inferred measure) is at least a weak argument for misperception.

It would be useful if a precise statement could be made about the magnitude and shape of the overconfidence function. Unfortunately, this is not yet possible except in a very crude and tentative way. It would appear that the function generally decreases with OP irrespective of how the uncertainty in the event is constituted or measured. However, there is always a very large individual difference component. Moreover, direct measures produce lower, and seemingly more valid, estimates than do choice-inferred measures, and some tasks seem to generate more overconfidence than others. Dart throwing has produced the greatest amount of overconfidence among the tasks studied to date. This could be attributed to the fact that of the tasks used so far, dart throwing is the most obviously devoid of external sources of control. Another problem in describing "the overconfidence function" is that certain OP levels, notably around .50, seem to have a distinctive effect on SP: people seem either to recognize "even" bets more readily than others, or to prefer the .50 response, for SP in this region shows a perceptible dip (although it remains above OP, see Fig. 1).

The second major conclusion has to do with *combination* of internal uncertainty values. It would appear that people are very nearly optimal in the rules they use to aggregate component uncertainties, at least in this present task. Given their initial overconfidence in each component event, and a relatively constant "adjustment" applied in the case of compounds, SP is accounted for quite nicely by a simple multiplicative rule. Consider, for example, the progression of obtained SP values for an extreme case: HH(.45), HHH(.27), and HHHH(.18). If we multiply each of these by .60 (the SP for H alone) to predict SP for the next higher compound, we get HHH(.27) and HHHH(.16), which is in close agreement with the values obtained. Thus the SP associated with H has a fairly constant and appropriate "impact" in all its compounds. This is not an isolated case, as is shown in Table 3.

Finally, a word is in order concerning the combination of internal and external uncertainty in the same compound event. This, after all, is probably most representative of real-life decision situations. In triple as in double compounds, people show the predicted preference for internal sources. However, for reasons yet to be determined, the magnitude of preference was not as strong here as in an earlier study. Moreover, the preference breaks down completely at very low OP levels where people show the greatest amount of overconfidence in internal sources and, presumably, in external sources as well. At this point, but certainly not at higher levels, it appears safe to regard SP from internal and external sources as strictly comparable.

REFERENCES

BEACH, L. R., & PHILLIPS, L. D. Subjective probability inferred from estimates and bets. *Journal of Experimental Psychology*, 1967, **75**, 354–359.

BEACH, L. R., & WISE, J. A. Subjective probability and decision strategy. *Journal of Experimental Psychology*, 1969, **79**, 133–138. (a)

BEACH, L. R., & WISE, J. A. Subjective probability revision and subsequent decisions. *Journal of Experimental Psychology*, 1969, **81**, 561–565. (b)

HOWELL, W. C. An evaluation of subjective probability in a visual discrimination task. *Journal of Experimental Psychology*, 1967, **75**, 479–486.

HOWELL, W. C. Intuitive "counting" and "tagging" in memory. *Journal of Experimental Psychology*, 1970, **85**, 210–215.

HOWELL, W. C. Uncertainty from internal and external sources: A clear case of overconfidence. *Journal of Experimental Psychology*, 1971, **89**, 240–243.

PETERSON, C. R., & BEACH, L. R. Man as an intuitive statistician. *Psychological Review*, 1967, **68**, 29–46.

ROTTER, J. B. Generalized expectancies for internal vs. external control of reinforcement. *Psychological Monographs*, 1966, **80** (1, Whole No. 609).

SLOVIC, P., & LICHTENSTEIN, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, 1971, **6**, 699–744.

WISE, J. H. Estimates and scaled judgments of subjective probabilities. *Organizational Behavior and Human Performance*, 1970, **1**, 85–92.

(Received August 7, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 14–19

EFFECT OF STIMULUS ORIENTATION UPON HAPTIC PERCEPTION OF THE HORIZONTAL–VERTICAL ILLUSION [1]

J. DEREGOWSKI [2] AND HADYN D. ELLIS

University of Aberdeen

The effect of orientation upon haptic perception of the L and T figures presented in *S*'s frontoparallel plane and in a horizontal plane in front of him was investigated. The data confirm the importance of the section effect and suggest that this effect is independent both of the inclination of the stimulus within a plane and of the plane within which it is presented. A theoretical curve showing the relationship between the radial and tagential components of haptic exploration was derived and found, prima facie, to provide a satisfactory approximation to the data.

Two recent, related studies (Day & Avery, 1970; Day & Wong, 1971) demonstrate certain disparities between visually and haptically perceived horizontal–vertical illusion figures. When an L figure is explored haptically, the perceived illusion is a result of a combination of the two types of exploratory movement described by Davidon and Cheng (1964). Those parts of the figure which evoke radial movements (i.e., movements along any of the infinity of lines (radii) intersecting at *S*) tend to be perceived as longer than those which evoke tangential movements (i.e., movements in planes to which any one of the radii is perpendicular.) In accord with these observations, it was found that when *S* is presented with an L-type stimulus in his frontoparallel plane (and, hence, is not required to make any radial movements), the illusion is absent. If an inverted-T figure is so presented, the usual illusory distortion of the figure occurs. Day and Avery attribute this distortion to the bisection of one line by another which is present in the T figure but not in the L figure. These results are, in part, at variance with the well-established findings obtained with geometrically similar visual stimuli presented in the same plane (Segall, Campbell, & Herskovits, 1966). In the case of the latter, the illusion occurs for *both* figures, and is, as here, greater for the T-type figure. This discrepancy in the magnitude of illusion has been investigated by Künnapas (1955), who also suggested that it was due to intersection of one line by another.

If the haptic section effect is the only factor distinguishing the T figure from the L figure, then one might expect this factor to remain constant whenever the two figures are compared under identical conditions. Should the difference in the amount of illusion evoked by the two figures under such conditions be shown to be variable, one would be forced to concede the existence of some other factor or factors in addition to simple section effect. Furthermore, when comparing such figures one would expect the magnitude of illusion observed to be a function of the radial and tangential components of exploratory movements performed by *S*.

The experiments reported below were intended to test the following hypothesis derived from the above considerations: the difference in the magnitude of haptic illusion between the inverted-T figure and an L figure when both figures are presented in the same plane and in the same orientation is independent of the orientation of these figures. They were also intended to explore the relationship between the two haptic components (radial and tangential) claimed to account largely for the illusion.

---

[1] We wish to acknowledge the assistance of Rosalind Heaton in collecting the data.

[2] Requests for reprints should be sent to J. Deregowski, Department of Psychology, University of Aberdeen, Aberdeen, Scotland.

EXPERIMENT I

Method

Apparatus.—The apparatus used was based on Day and Avery's (1970) design with a single, but important, modification, which allowed the orientation of the figure to be varied. The apparatus consisted of a board with slots into which strips of polished metal could be firmly set to form either a T or an L figure, suitable for haptic exploration. The metal strips used were about 1.5 mm. thick and when set, edgewise, in place, protruded about 6 mm. above the face of the board. The entire arrangement was set within a large turntable, which could be rotated in the plane of the board. A scale attached to the edge of the turntable enabled *E* to measure the extent of its rotation. A set of 25 metal strips was used, varying in length, in 2.5-mm. steps, from 40 mm. to 100 mm., as well as a standard strip 75 mm. long. The turntable could be placed either in the horizontal or in the vertical plane.

Subjects.—There were 24 volunteer *S*s (12 men and 12 women) who were paid for participation and drawn from students (both undergraduate and postgraduate) and staff (secretarial, ancillary and teaching) of the University of Aberdeen.

Procedure.—The same procedure was used with both figures. Each *S* was asked to sit at the table on which the apparatus was placed with the turntable in the horizontal plane and to put on a pair of translucent goggles. The shield covering the metal strips was then removed and he was told that he would be required to explore a shape with the middle finger of his preferred hand and to decide whether one of the lines forming the shape is longer or shorter than the other. He was asked which was his preferred hand.

The *E* took *S*'s preferred hand and placed it on the apparatus with the middle finger in contact with the junction point of the two lines. Choosing the initial direction at random, *S*'s hand was guided over the entire shape and he was asked, whilst exploring the variable line, whether "this line" was longer or shorter than "this [the standard] line." (At this point, *S*'s hand was guided to the standard line and traversed it twice.)

After this introductory exploration, *S*'s hand was released and he was told that he now was allowed to make six exploratory traverses of his own before making up his mind.

This procedure was repeated for each of the stimuli presented to *S*. If the need arose, in the course of the experiment, *S*s were reminded that they had to use the middle finger only in exploring the figure.

Each *S* responded to one set of stimuli to which he was allocated at random, with a restriction that of four *S*s allocated to each set two were men and two were women. Each set was made of four stimuli: an inverted-T figure and an L figure in traditional (0°, datum setting) orientation, and of both figures in one of the following settings, 15°, 30°, 45°, 60°, 75°, 90°, the angle being measured clockwise from the datum setting at one of the other angles. The datum setting was introduced to provide base lines against which each *S*'s responses to other settings could be measured. When a set was being administered the stimuli were presented in a random order, with the restriction that neither two L stimuli nor two T stimuli followed each other in the sequence of trials. The constant-length limb of the L figure was either in *S*'s frontoparallel plane or occupied one of the positions in the right-hand quadrant nearer to *S* (see Fig. 1). For the inverted-T figure, the length of the bisected member was likewise kept constant and the entire figure rotated so that the bisecting member lay in various settings in the right-hand quadrant further from *S*.

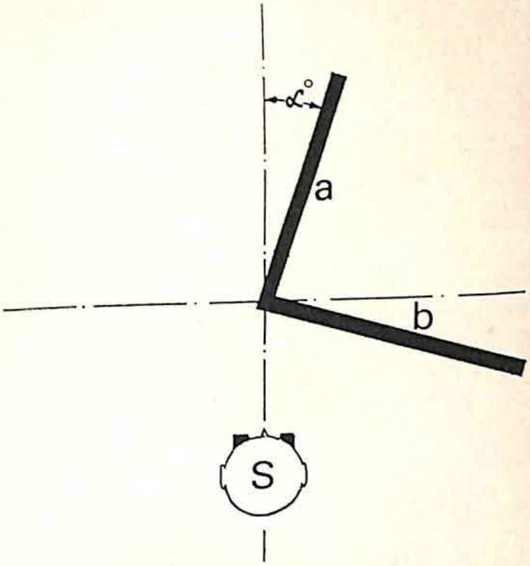


FIG. 1. An L figure at an inclination as presented in the first experiment. (Under these conditions the radial components of the two limbs are $a\cos\alpha$ and $b\sin\alpha$. These components are assumed to be subject to overestimation (factor K) so that the perceived length of the two limbs are $a(1 + K\cos\alpha)$ and $b(1 + K\sin\alpha)$, respectively. Hence determination of K using one of the PSE values yields an equation predicting the length of the adjustable limb (a) at all other settings within the quadrant: $a = b(1 + K\sin\alpha)/(1 + K\cos\alpha)$. This deduced curve is plotted in Fig. 2 (Curve D).)

The points of subjective equality (PSEs) perceived by *S* were determined by means of a staircase method (Cornsweet, 1962). A single staircase was used, in order to reduce decrements caused by practice (cf. Over, 1967). Coarse 5-mm. steps were initially used till three reversals were reached. After the third reversal, 12 trials, using 2.5-mm. steps, were run. The average of these 12 trials was taken to locate the PSE. Within each group of four *S*s responding to the same set of stimuli, one of the *S*s of each sex was tested by the ascending staircase method; the other by the descending method.

Results

The PSEs obtained with the two datum figures (inverted T and L) were compared using analysis of variance. The only significant factor was the nature of the stimuli, $F\ (1,\ 36) = 9.6,\ p < .005$. As expected the illusion was larger with the inverted-T stimulus (PSE = 58.1 mm., $s^2 = 61.6$) than with the L stimulus (PSE = 65.7 mm., $s^2 = 77.8$). A comparison of the PSE obtained with the L figure with the length of the standard (75.0 mm.) also yielded a significant result, $t(23) = 3.2,\ p < .005$. It therefore appears that both figures evoked an illusion, and that the illusion was 7.6 mm. (10.1%) larger in the case of the T figure.

The product-moment correlation between the PSEs obtained with the two datum figures was .53 ($p < .01$).

In order to reduce inter-*S* variation and to obtain a measure of illusion due to orientation only, PSEs obtained from each *S* at his nondatum orientation were divided by his PSE obtained with the same figure at the datum setting. The ratios thus obtained were used in an analysis of variance. Only one factor was found to be significant: the orientation of the figure, which was tested between *S*s, $F\,(5,\,18) = 3.06,\ p < .05$. Both effects tested within *S*s, i.e., the type of figure and the Type of Figure × Orientation interaction yielded $F < 1$.

Figure 2 shows PSE values for both stimuli at various orientations.

Discussion

The data obtained are in agreement with the results reported by Day and Wong (1971). Haptic illusion was evoked by both L and inverted-T figures when these were so placed that they called for exploratory movements in the horizontal plane. The consistently larger illusion obtained with the inverted-T figure

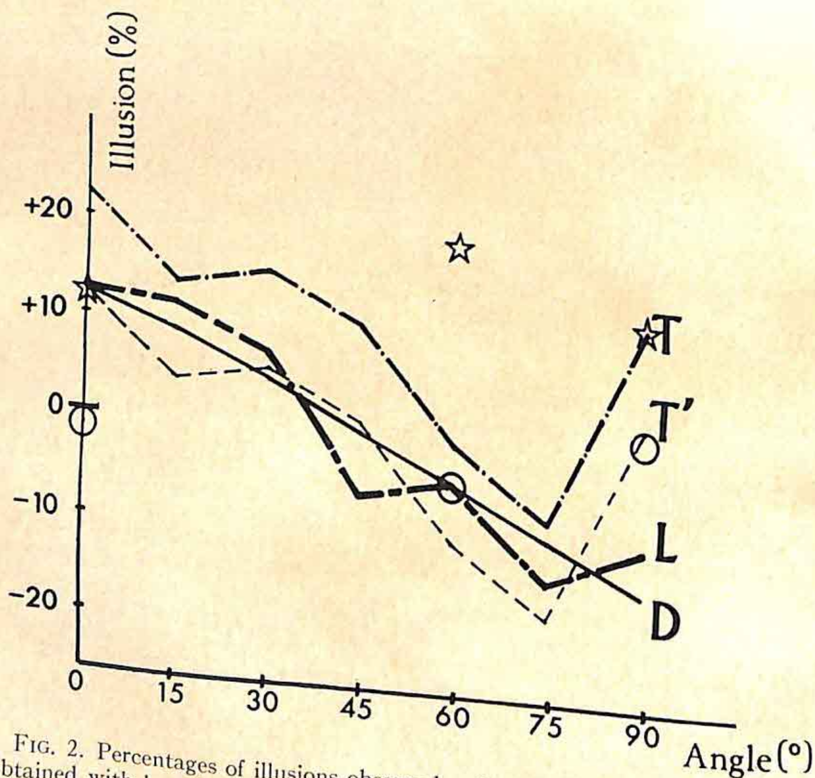


FIG. 2. Percentages of illusions observed. (Curves L and T represent data obtained with L and inverted-T figures in the first experiment. Curve D is a theoretically deduced curve as described in Fig. 1. Curve T′ is the T curve shifted to take account of the section effect. Open circles represent L-figure data and open stars represent the inverted-T-figure data obtained in the second experiment.)

agrees with earlier observations (Day & Avery, 1970) of the importance of the section effect. The nonsignificance of the within-*S*s comparison, of ratios of nondatum response to datum response for the two figures, shows that the present data provide no indication that this effect is dependent upon the orientation of the T figure. The significance of the product-moment correlation, on the other hand, suggests that a common mechanism contributes to the illusory effect in both figures.

The orientation of the stimuli in the horizontal plane, which was shown to have a significant effect upon illusion, affects the distribution of radial and tangential components of exploratory movements made when traversing the figures. It is, therefore, of interest to consider the present findings in relation to earlier observations, involving these components.

Cheng (1968) investigated the effect of the radial and the tangential components upon perception of distances between stimuli placed outwith the frontoparallel and the sagittal planes, and found that his results confirmed the earlier findings; *S*s tended to perceive the radial distances as longer than the tangential distances. However, in this, as in the earlier study (Davidon & Cheng, 1964), only purely radial and purely tangential arrangements were used. These form rather a special case, for most lines are neither radial nor tangential to *S* but contain both these components. This is the case with the horizontal–vertical illusion figures presented, as here, in skew orientations. In order to relate the present findings to the radial-tangential framework the following approach was used.

It was assumed that the radial element is always overestimated by a constant factor K and that the perceived length of a line is a sum of its true length and its illusory radial overestimation. Thus, in the case of an L-type figure presented to S at angle α to the sagittal plane in the quarter under consideration (Fig. 1), the *perceived* lengths are: $a(1 + K \cos \alpha)$ and $b(1 + K \sin \alpha)$, where a and b are the true lengths of the two limbs. Therefore, at the point of illusory equality of the two limbs, the ratio lengths of the two limbs are given by $a/b = (1 + K \sin \alpha)/(1 + K \cos \alpha)$. Given the values of a and b at any angle α the value of K can be determined and used in predicting values of illusion at other angles. Since (as it is a mean of 24 PSEs) the most reliable estimates of the a/b ratio obtained in the present experiment is that at the datum setting ($\alpha = 0°$), this value was used in deriving the theoretical curve for the illusion (Fig. 2). This curve should, if the section component of the inverted-T figure has a constant value, describe the L illusion and be parallel to the curve applicable to the inverted-T figure. This becomes apparent if one considers an analogous formula for the perceived length of the bisecting member of the T figure: $(a + c)(1 + K \cos \alpha)$, where c is the section effect. The section effect can, if Day and Wong (1971) are correct in suggesting that it is the only effect when no radial movements are called for, be determined by measuring the magnitude of illusion evoked by a T figure presented in any position in the frontoparallel plane. This was done in the following experiment.

EXPERIMENT II

Method

Apparatus.—The same apparatus as in Exp. I was used.

Subjects.—There were eight volunteer *S*s (four of each sex) paid for participation. Seven were drawn from staff of the University (secretarial and teaching) and the eighth from the staff of a secondary school.

Procedure.—The procedure was similar to that used previously. Here, however, the figures were presented to *S*s in their frontoparallel planes and apart from the datum setting, in which one limb of the L figure and the stem of the inverted-T figure were vertical, only two settings were used. These were 60° and 90°, the angles, as seen by the *S*s, being measured clockwise from the datum setting. These angles were suggested by the inspection of the T-figure graph obtained in Exp. I, which showed no illusion at $\alpha = 60°$ and a surprisingly large illusion at $\alpha = 90°$. The *S*s were divided into two groups. Both groups responded to datum figures, and, in addition, one of them responded to figures presented at $\alpha = 60°$, the other to figures presented at $\alpha = 90°$. The order of presentation of figures and the allocation of *S*s to the two groups were governed by the same criteria as in Exp. I.

Results

The mean values of PSEs obtained are plotted in Fig. 2. A comparison between datum responses to the two figures revealed as expected a larger illusion in respone to the T figure, $t\ (7) = 3.33\ p < .01$. An analysis of variance of the responses made to the inclined figures yielded only one significant effect; this too was the difference between the two types of figures, $F\ (1,\ 6) = 29.7,\ p < .01$, the T figures yielding larger illusion. Neither the angle of inclination nor the Angle $\times$ Type of Figure interaction reached significance.

A comparison of the mean PSE of the L-figure mean at datum setting, (77.5 mm., $s^2 = 82.8$) with the no-illusion value of 75 mm., showed no significant difference, $t(7) = .78$, $p > .2$. A product-moment correlation of responses to the two figures made by all *S*s was found to be nonsignificant, $r(6) = .48$, $p > .10$.

If the illusion observed with the T figure is primarily due to the section effect, one would expect no illusion to be observable in the case of the L figure. Furthermore, one would expect the T- and L-figure responses elicited from the same *S* to be uncorrelated. The data obtained are in accord with the above expectations and can, therefore, be said to support the conclusions of Day and Wong (1971).

An additional support for these conclusions is provided by the following consideration. The data obtained under all three conditions in which the T figure was used served in determining the magnitude of the intercept *c* separating the L and T curves. This was derived by subtracting the overall mean PSE obtained with the T figure (65.30 mm., $s^2 = 93.1$) from the length of the standard (75.00 mm.), This gave $c = 9.7$ mm. Using this value, an adjustment to the T-figure data obtained in Exp. I was carried out (Fig. 2, curve T′) and these data were compared with the corresponding data obtained with the L figure. No significant difference was found, $t(23) = 1.26$, $p > .10$.

Discussion

Both experiments appear to support earlier findings of Day and Avery (1970) and Day and Wong (1971), and offer a suggestion for the possible extension of the radial-tangential analysis (Cheng, 1968; Davidon & Cheng, 1964) to figures which do not present purely radial-tangential orientations. Inspection of the theoretical curve for the L figure and its relationship to the empirical points, other than the single point needed to obtain the value of the constant *K*, suggests that a plausible explanation has been offered by the earlier workers, which is more general in scope than the data on which it was based would imply. Similarly the inverted-T figure data, when the section effect is removed, fall, with the sole exception of the PSE obtained with the figure at $\alpha = 90°$, within proximity of the theoretical curve. The discrepancy at 90° merits further investigation; for it may be that this orientation calls forth some special and hitherto unconsidered factor of the perceptual mechanism.

Since the apparatus and the procedure used are similar to those of Day and Avery (1970) and of Day and Wong (1971), it is of interest to compare the mean PSEs obtained by these workers with those obtained here with identically oriented stimuli.

For the L figure presented in the frontoparallel plane, the following mean PSEs were obtained: 77.5 ($s^2 = 82.8$) for the present data; 76.1 ($s^2 = 35.0$) by Day and Avery (1970), and 74.6 ($s^2 = 36.75$) by Day and Wong (1971). For the same figure in the horizontal plane the values were 65.7 ($s^2 = 78.00$) for the present data and 60.6 ($s^2 = 57.76$) by Day and Wong (1971). Similarly for the T figure presented in the frontoparallel plane the PSEs were 66.1 ($s^2 = 97.80$) and 62.6 ($s^2 = 40.00$) by Day and Avery (1970). It appears, therefore, that the mean values obtained in the present experiments closely approximate the previously reported results, and can probably be regarded as confirmatory of these findings. These values also corroborate the suggestion (Day and Wong, 1971) that there is an essential difference between the visual and the haptic perceptual mechanisms since the L figure is capable of evoking a visual illusion whether in the horizontal or the vertical plane, whereas when explored haptically the L figure evokes illusion only if placed horizontally. Furthermore, when placed in the frontoparallel plane, the T figure appears to evoke an illusion due to the section effect only, contrary to the results obtained by Künnapas (1955) in the visual modality where-in both the direction of the lines and the section effect influenced perception. Since, however, the section effect is present in both modalities it may be that the mechanisms, although not identical, share this characteristic; but, whereas in the case of visual stimuli the illusory effect not due to the section is a function of retinal meridians (Avery & Day, 1969), in the case of tactile stimuli it can be described in terms of radial and tangential components of exploratory movements.

REFERENCES

Avery, G. C., & Day, R. H. Basis of the horizontal-vertical illusion. *Journal of Experimental Psychology*, 1969, **81**, 376–380.

Cheng, M. F. Tactile-kinaesthetic perception of length. *American Journal of Psychology*, 1968, **81**, 74–82.

Cornsweet, T. The staircase method in psychophysics. *American Journal of Psychology*, 1962, **75**, 485–491.

Davidon, R. S., & Cheng, M. F. Apparent distance in a horizontal plane with tactile-kinaesthetic stimuli. *Quarterly Journal of Experimental Psychology*, 1964, **16**, 277–281.

Day, R. H., & Avery, G. C. Absence of the horizontal–vertical illusion in haptic space. *Journal of Experimental Psychology*, 1970, 172–173.

Day, R. H., & Wong, T. S. Radial and tangen movement directions as determinants of the hap illusion in an L-figure. *Journal of Experime Psychology*, 1971, **87**, 19–22.

Künnapas, T. M. An analysis of the "vertic horizontal illusion." *Journal of Experime Psychology*, 1955, **49**, 134–140.

Over, R. Intermanual transfer of practice crements with a haptic illusion. *Quarterly Jour of Experimental Psychology*, 1967, **19**, 215–218.

Segall, M. H., Campbell, D. T., & Herskovi M. J. *The influence of culture on visual percepti* Indianapolis: Bobbs-Merrill, 1966.

(Received August 12, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 20–24

THE HUMAN BODY AS STIMULUS OBJECT: ESTIMATES OF DISTANCES BETWEEN BODY LANDMARKS

FRANKLIN C. SHONTZ[1] AND RONALD D. McNISH[2]

University of Kansas

The *S*s were exposed to varying degrees of objectification of the human body to test the hypothesis that features characteristic of responses to body stimuli are the product of direct body experience. Experiment I demonstrated that judgment of a mirror image of the self produces responses similar to those produced when the body itself is the stimulus. Experiment II further objectified body stimuli to determine at what point responses become similar to those produced by nonbody objects. Features characteristic of responses to body stimuli did not appear when *S* could not recognize the stimulus as a human figure. Results support the revised hypothesis that stimuli recognized as human figures produce different perceptual responses than stimuli not recognizable as human figures.

Several investigators have suggested that responses to stimuli on the body differ from responses to objects in environmental space (Fisher, 1970; Fisher & Cleveland, 1968; Shontz, 1969; Wapner & Werner, 1965). The source of these differences has not been clearly identified, but investigators have generally assumed that central psychological processes, variously labeled *body schemata*, *body image*, and *body concept*, are responsible.

Shontz (1969) found a stable pattern of error scores in *S*s' estimates of certain distances on their own bodies: head width and forearm length were consistently more overestimated (or less underestimated) than hand length and foot length. As compared to estimates of distances in environmental space, responses to body distances were more variable and less accurate overall. Several of these findings have been independently confirmed by Fuhrer and Cowen (1967) and Hester (1970).

The following two experiments attempted to objectify the human body, that is, to make it more like an object in environmental space. The purpose of objectification was to test the hypothesis that characteristic features of responses to body stimuli are the product of direct body experience. If that hypothesis is correct, these features will appear in responses to stimuli on the personal body but not in responses to other stimuli, however similar to the personal body they may be.

EXPERIMENT I

Method.—The *S*s were six groups of three men and three women each from an introductory course in psychology, in which participation as *S*s in psychological research was required. Each *S* estimated five body distances and the lengths of five nonbody stimuli under three experimental conditions. The body distances were selected from a longer list of stimuli used in previous research (Shontz, 1969, pp. 100–101). The distances employed in this investigation were as follows:

1. Head width—which was defined as the width of your head across the temples, from a point just above the tip of one ear on one side to a similar point just above the tip of the ear on the opposite side. (Judge the straight-line distance between these points.)
2. Forearm length—which was defined as the distance from the point of your right elbow to the center of the bone bump at your right wrist.
3. Waist width—which was defined as the width of your waist, from a point at your belt line on the right side of your body straight through to a similar point on your belt line on the left side.
4. Hand length—which was defined as the distance from the point where your right wrist joins your hand to the tip of your longest finger, your middle finger. (Judge the distance when your hand is outstretched.)
5. Foot length—which was defined as the length of your right foot, from the back of your heel to the tip of your longest toe.

Average actual body distances of male *S*s in Exp. I were: head width, 15.2 cm. (*SD* = .6); forearm length, 27.0 cm. (*SD* = 1.6); waist width, 26.7 cm. (*SD* = 1.7); hand length, 19.8 cm. (*SD* = 1.4); foot length, 26.4 cm. (*SD* = 1.5). Average actual body distances of female *S*s in Exp. I were head width, 14.8 cm. (*SD* = .6); forearm length, 26.8 cm. (*SD* = 1.3); waist width, 23.3 cm. (*SD* = 1.3); hand length, 18.2 cm. (*SD* = .9); foot length 17.3 cm. (*SD* = 1.0).

---

[1] Requests for reprints should be sent to Franklin C. Shontz, Department of Psychology, University of Kansas, Lawrence, Kansas 66044.

[2] The second author was supported in part by National Science Foundation Undergraduate Research Participant Grant GY 8921.

The nonbody stimuli were 1.3-cm. diameter rods, 12.6, 28.6, 38.5, 43.3, and 59.9 cm. long.

Body and nonbody stimuli were presented alternately, although a nonbody stimulus was always presented first. Otherwise, order of stimulus presentation was random and differed from *S* to *S* as well as from condition to condition.

The three experimental conditions were: Standard; Mirror, Object-Oriented; and Mirror, Self-Oriented. Under Standard conditions, *S* was seated in front of the response apparatus and his body was covered by a drape so that he could see directly neither his own body parts nor the nonbody stimuli. Under the Mirror, Object-Oriented condition, *S* viewed himself and the nonbody stimuli in a 1.52-m., square mirror, 1.83 m. to his left and parallel to his mesial plane (body drape in place). The *S* was instructed to estimate appropriate distances as seen in the mirror. Under the Mirror, Self-Oriented condition, *S* was instructed to consult the mirror image of his own body for purposes of stimulus identification but to judge the distances on his own body as he currently experienced it.

The *S*s were randomly assigned to groups, and the three experimental conditions were presented in different orders to the six groups. Body distances were specified by verbal description. Rods were handed to *S*s under the body drape.

The *S*s provided estimates on an apparatus consisting of a 3.2-cm.-diameter, 1.83-m.-long wooden rod, mounted horizontally, 24.1 cm. above a table, which was 71.1 cm. high, 1.83 m. wide, and 53.3 cm. deep. Pointers on the rod were attached to the movable elements of a traverse rod and were manipulated by *S*, who pulled ropes extending from the pointers to his hands, under the body drape. Pulling one rope separated the pointers; pulling the other rope brought them closer together. The *S* was free to move the pointers back and forth until satisfied with the setting of the pointers; *E* returned the pointers to a closed position in the center of the rod after recording each response. A screen prevented *S* from seeing *E* while data were recorded. Previous research had established that response characteristics are not affected by *S*'s use of his hands to adjust the pointers and that the procedure employed in this experiment provides reliable data (Shontz, 1969).

Because distance on the personal body vary among *S*s, response magnitudes were interpretable only in relation to measures of individual body distances. The basic data were therefore percentage scores, each response expressed as a percentage of stimulus length. Statistical analyses were also conducted on raw error scores (response size minus stimulus size) with adjustment for variations in stimulus sizes by analysis of covariance. Because outcome was unaffected by mode of analysis, and because percentage scores have been used in previous research, only analyses based on percentage scores are reported below.

TABLE 1

PERCENTAGE SCORES FOR BODY DISTANCES IN EXPERIMENT I COMPARED TO SIMILAR DATA FROM PREVIOUS RESEARCH

Body part	Exp. I		Previous	
	M	*SD*	*M*	*SD*
Head	127.04	17.9	121.8	17.6
Forearm	118.01	13.3	114.5	20.1
Waist	126.32	12.7	116.2	17.0
Foot	97.20	13.8	96.6	11.5
Hand	100.69	13.5	93.3	10.4

Results.—In the complex analysis of variance that resulted from the design of Exp. I, only two terms were significant: conditions, F (2, 48) = 22.44, $p < .01$, and stimuli, F (9, 216) = 23.54, $p < .01$. The Mirror, Object-Oriented condition yielded lower percentage scores overall ($\bar{X}$ = 100.7; *SD* = 22.9) than did the Standard ($\bar{X}$ = 105.1, *SD* = 21.6) or Mirror, Self-Oriented ($\bar{X}$ = 106.4; *SD* = 22.7) conditions. Sex differences and differences among orders of administration of stimuli were not significant, either as main effects or in interactions with other terms in the analysis.

The difference between the mean for combined body stimuli (113.8; *SD* = 14.7) and the mean for combined nonbody stimuli (94.3; *SD* = 23.9) accounted for 48.3% of the total variance among stimuli. Variation among means for body stimuli accounted for 40.2% of the total variance among stimuli. Variation among means for nonbody stimuli accounted for only 11.4% of the total variance among stimuli. The distribution of means for body stimuli, shown in Table 1, follows the general pattern described previously by Shontz (1969, p. 236). Head width and forearm length were relatively overestimated as compared to hand and foot lengths. (Waist width is included because of its possible interest as a source of sex differences. See Shontz, 1969.)

Results were unchanged when data from only the condition presented first to each group were analyzed. The findings are therefore not a function of the use of repeated measures in the research design.

Discussion.—The first experiment disconfirmed the hypothesis. Errors of estimate of distances on the body displayed similar characteristics, whether *S*s judged body distances directly or saw their bodies as objects in a mirror. The sample of body stimuli was overestimated in comparison with the sample of nonbody stimuli. Variability of percentage scores was greater among body stimuli than among nonbody stimuli, and error patterns were consistent with those found in previous research.

The overall reduction of error scores for combined body and nonbody stimuli under the Mirror, Object-Oriented condition implies that experimental conditions affected body and nonbody stimuli similarly. This finding therefore does not support the hypothesis that mode of stimulus presentation exerts different effects on stimuli of these two types.

The experiment did not show whether attempts to further objectify body stimuli would alter this pattern of results or at what point in a succession of stimuli, each progressively different from the personal body, the characteristic pattern would fail to appear.

In a pilot study, one of the *E*s (R. M.) asked *S* to estimate distances on *E*'s body. The same pattern of percentage scores appeared in *S*'s responses to this stimulus as appeared in estimates of distances on his own body; Exp. II was designed to determine the limits of this phenomenon.

EXPERIMENT II

Method.—Sixteen groups of two men and two women each were drawn in a subsequent semester from the pool of *S*s described above. The *S*s were randomly assigned to groups. Each *S* responded to 13 stimuli under one of eight experimental conditions. Five stimuli were personal body distances administered in a manner similar to the Mirror, Self-Oriented condition of Exp. I; 3 stimuli were wooden rods, 14.6, 21.7, and 28.6 cm. long. These 8 stimuli were presented to all *S*s.

The other five stimuli were either distances on other persons' bodies or drawings constructed to represent distances on the human body. Variations among these stimuli defined the eight experimental conditions. In the Same-Sex, Direct-Observation condition, *S* judged distances on the body of a person of the same sex as himself. In the Other-Sex, Direct-Observation condition, *S* judged distances on the body of a person of the opposite sex. In the Same-Sex, Mirror-Image and Other-Sex, Mirror-Image conditions, *S* viewed an appropriate stimulus person in the mirror used in Exp. I. To fulfill these requirements, *S*s were brought to the laboratory in pairs and served as stimuli for each other as well as respondents. Persons whose body distances were judged sat in an armless, straight-backed chair, 1.83 m. from *S*; rigid posture was not required, but gross movement of body parts was not permitted.

In the Same-Sex, Outline-Drawing condition, *S* estimated distances on a full size outline drawing of a standing figure of the same sex as the *S*. In the Other-Sex, Outline-Drawing condition, *S* estimated distances on a full size outline drawing of a standing figure of the opposite sex. In the Abstract Figure condition, *S* estimated distances on a robot-like drawing with a contour of straight lines, but having only vaguely human characteristics (Fig. 1). In the Dots condition, *S* estimated distances between chalk marks identified by letters of the alphabet. The dots were located at positions identified by letters in Fig. 1, but connecting lines were omitted from the drawing. All drawings were prepared on portable chalkboards, removed from their stands, and placed on edge before the covered mirror. Distances corresponding to head width, forearm length, waist width, hand length, and foot length were similar in all drawings (Table 2) and approximated average distances on *S*s' own bodies. Average actual body distances of male *S*s in Exp. II were: head width, 15.5 cm. ($SD = .5$); forearm length, 26.9 cm. ($SD = 1.2$); waist width, 26.6 cm. ($SD = 2.5$); hand length, 19.0 cm. ($SD = 1.0$); foot length, 26.0 cm. ($SD = 1.4$). Average actual body distances for female *S*s in Exp. II were head width, 14.9 cm. ($SD = .5$); forearm length, 24.6 cm. ($SD = 1.2$); waist width, 22.5 cm. ($SD = 1.7$); hand length, 17.8 cm. ($SD = 1.0$); foot length, 23.1 cm. ($SD = 1.2$).

Eight groups estimated personal body distances first; eight groups estimated experimental stimuli (distances on drawings or on other persons' bodies) first. Nonbody stimuli were administered in the middle of the stimulus series to all *S*s. Basic data were the same as those in Exp. I.

Results.—Sex differences and differences between orders of administration of stimuli were not significant, either as main effects or in interaction with other terms in the analysis. Terms significant at the .01 level in the analysis of variance were stimuli, $F\ (12, 384) = 50.20$, and the Stimuli × Conditions interaction, $F\ (84, 384) = 1.72$. The latter is critical to the interpretation of the findings.

To determine the influence of experimental conditions on percentage scores, the variance among scores within each condition was analyzed into three components. The first component was the variance accounted for by differences

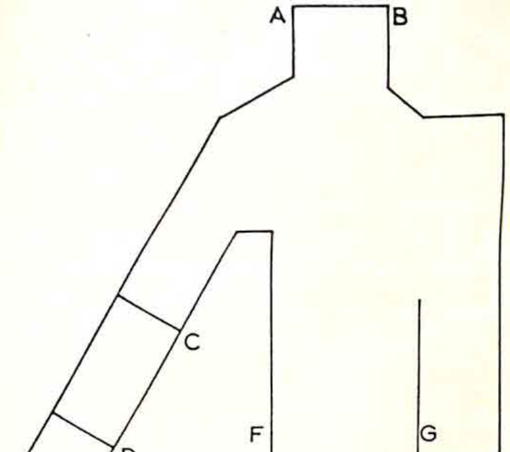


FIG. 1. Abstract figure used in Exp. II.

TABLE 2

STIMULUS DISTANCES (IN CM.) ON PREPARED DRAWINGS IN EXPERIMENT II

Distance	Outline (male)	Outline (female)	Abstract figure	Unconnected dots
Head width	16.6	14.7	16.7	16.7
Forearm length	24.6	23.6	21.8	21.8
Waist width	25.4	20.1	24.7	24.7
Hand length	19.3	17.4	17.7	17.7
Foot length	24.8	20.7	25.0	25.0

among scores for personal body parts. The second component was the variance accounted for by differences among scores for stimuli that defined each of the eight experimental conditions. The third component was the variance accounted for by differences among scores for the three wooden rods, the lengths of which were judged by all *S*s in all experimental conditions. Inspection of the data showed that it was feasible to combine the eight experimental conditions (second component) into three groups: other bodies (both direct and in the mirror); drawings of bodies; and nonrepresentational drawings (abstract and dots figures).

Where responses to other bodies or to representations of other bodies parallel responses to the personal body, variances among stimuli are expected to be relatively high. The first column of data in Table 3 shows that the percentages of variance accounted for by differences among scores for stimuli on the personal body is uniformly large. The second column of data shows that the percentages of variance accounted for by differences among stimuli within experimental conditions drops markedly as conditions become more objectified (less recognizable as a human body). The third column shows that the variance accounted for by the wooden rods is uniformly low.

With this sample of stimuli, average percentage scores of responses to distances on the body are expected to exceed average percentage scores of responses to nonbody objects. Table 4 shows that average percentage scores decreased and became more accurate (approached the value of 100) as stimuli became more objectified.

To establish the extent to which percentage scores conformed to the expected pattern under each condition, index values (Shontz, 1969, p. 185) were calculated and compared. The sum of the percentage scores for hand length and foot length was subtracted from the sum of the percentage scores for head width and forearm length. A high positive value indicates conformity

TABLE 3

PERCENTAGES OF VARIANCES ACCOUNTED FOR BY DIFFERENCES AMONG STIMULI OF DIFFERENT TYPES IN EXPERIMENT II

Experimental cond.	Personal body	Other stimuli	Wooden rods
Other bodies	44.0	43.6	1.4
Drawings of bodies	58.2	31.3	.4
Nonrepresentational drawings	52.6	11.4	3.9

Note.—Personal body distances and lengths of wooden rods were judged by all *S*s in all conditions.

TABLE 4

Mean Percentage Scores for Body Distances and Other Stimuli in Experiment II

Statistic	Personal body	Others' bodies	Drawings of bodies	Nonrepresentational drawings	Wooden rods
M	114.5	111.5	111.7	103.6	101.6
SD	21.5	17.4	23.7	21.8	13.0
N	64	32	16	16	64

Note.—Stimulus types are arranged from left to right in order of increasing objectification. Personal body distances and wooden rods were judged by all *S*s. Stimuli in other columns were judged by *S*s assigned to particular experimental groups.

of results to the expected pattern. Table 5 shows that differences between average index values for personal body distances and for experimental (other) stimuli increased as stimuli became less like the human body. The difference was statistically significant for nonrepresentational drawings, F (1, 15) = 23.26, $p < .01$.

CONCLUSIONS

The two experiments disconfirm the hypothesis that response characteristics, typical of estimates of distances on the personal body, occur only as a result of direct body experience. They also appear in estimates of distances on mirror images of the personal body, in estimates of distances on the bodies of others (viewed directly or in the mirror) and in estimates of distances on two-dimensional drawings of the human figure.

Location of stimuli in space provides an unlikely explanation of these results. The *S*s viewed themselves and other persons in a sitting position, but drawings of bodies were presented with the figure in a standing position, yet results were quite similar for both types of stimuli. Furthermore, the points on the nonrepresentational figures corresponded in spatial location to body stimuli, but responses typical of those to body stimuli did not appear when these configurations were administered. Contour alone also provides an unlikely explanation, since the abstract figure was contoured, but was not responded to as were bodies or obvious representations of bodies.

TABLE 5

Index Values for Personal Body Distances and Other Stimuli

Experimental cond.	Personal body		Other stimuli	
	M	*SD*	*M*	*SD*
Other bodies	46.06	28.8	49.20	20.8
Drawings of bodies	54.25	22.7	42.44	27.4
Nonrepresentational drawings[a]	48.96	24.6	21.49	19.7

Note.—Index value = (head width + forearm length) − (hand length + foot length).

[a] Difference between means significant at $p < .01$.

Recognition of the stimulus as a human body does seem important. After data were collected, appropriate *S*s were asked whether the abstract figure looked like anything recognizable to them. Only two out of eight who saw the figure mentioned a resemblance to a human body; both remarked that the resemblance was weak. Neither had noted any similarity of the figure to a human body while providing distance estimates, and data from these two *S*s were in no way remarkable or different from data provided by others in the same group. The rest of the *S*s identified the figure as a building or said it had no meaning to them.

The results are consistent with the revised hypothesis that recognition of a stimulus as a human body activates cognitive processes (probably derived from personal body experience) and produces perceptual responses that differ systematically from those that occur to stimuli which are not so recognized. Disconfirmation of this hypothesis requires demonstrating that characteristics of responses to recognized body stimuli also occur when stimuli possess important physical features of the human body but are not interpreted as being a body by *S*. In Exp. II, these conditions failed to produce response patterns similar to those produced by readily recognized body configurations.

REFERENCES

Fisher, S. *Body experience in fantasy and behavior.* New York: Appleton-Century-Crofts, 1970.

Fisher, S., & Cleveland, S. E. *Body image and personality* (2nd rev. ed.). New York: Dover, 1968.

Fuhrer, M. J., & Cowen, C. O. Influence of active movements, illumination, and sex on estimates of body-part size. *Perceptual and Motor Skills,* 1967, **24**, 979–985.

Hester, G. A. Effects of active movement on body-part estimates. *Perceptual and Motor Skills,* 1970, **30**, 607–613.

Shontz, F. C. *Perceptual and cognitive aspects of body experience.* New York: Academic Press, 1969.

Wapner, S., & Werner, H. (Eds.) *The body percept.* New York: Random House, 1965.

(Received August 14, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 25–31

HEMISPHERIC ASYMMETRY:
VERBAL AND SPATIAL ENCODING OF VISUAL STIMULI[1]

GINA GEFFEN,[2] JOHN L. BRADSHAW, AND NORMAN C. NETTLETON

Monash University, Australia

Differences in response time between physical and name matches were studied while investigating hemispheric specialization. Stimuli which had the same name were responded to more quickly when they were presented in the right visual field, whereas matching was faster when physically identical stimuli were presented in the left visual field. These differences could reflect interhemispheric transfer of information or asymmetry between the hemispheres in their processing capacities for different tasks. The effects of varying expectancies were found to be minimal: visual field differences in response time changed little whether *S* knew in advance, or did not know, which hemifield would be stimulated. The advantage of the left visual field in processing physical matches was greater when *S*s did not know whether a name or physical match would be required.

Posner and Mitchell (1967) have studied visual and verbal encoding of stimuli in a same–different reaction time (RT) task and shown that when a pair of stimuli are physically identical (e.g., AA), the response is about 70 msec. faster than when the stimuli share the same name but are printed in different cases (e.g., Aa). The projection of stimuli which require a name match (Aa) to the right visual field (RVF), and hence to the left (language) hemisphere, would be expected to produce faster RTs than when the same stimuli are projected to the left visual field (LVF) and hence to the right hemisphere. Geffen, Bradshaw, and Wallace (1971) have found that nonverbal stimuli, such as faces, were processed faster when presented for matching to the LVF. Conversely, verbal stimuli, which required identification, were processed more quickly when presented in the RVF.

Similar results have been obtained in a memory-scanning task by Klatzky (1970) and Klatzky and Atkinson (1971). They reported faster processing by the right hemisphere of test stimuli requiring a spatial match and, conversely, left hemisphere superiority for test stimuli which required verbal processing. One difficulty in interpreting their findings is in deciding whether the memory set is stored in the left or right hemisphere, or in both. Furthermore they used a 400-msec. exposure interval for the test stimulus, which makes its hemispheric location difficult to define. It would seem necessary to minimize the contribution of both eye movements and memory in order to investigate perceptual asymmetry. Egeth (1971) found that with unilateral presentation of circles and squares in a same–different RT task there was no visual field difference. However, half of his *S*s used their left hands and showed a 19-msec. difference in favor of the LVF, whereas the group who used their right hand showed a 1-msec. difference in the opposite direction. On repeating the experiment with interspersed trials of bilaterally presented stimuli, Egeth subsequently found that within trials involving unilateral presentation, stimulus pairs presented to the LVF were processed significantly faster, by 60 msec., and interpreted these results in terms of implicit postexposural scanning. However, both sets of results would appear to support right hemisphere superiority for processing nonverbal, visuospatial material.

Right hemisphere superiority for nonverbal stimuli has proved more elusive than left hemisphere dominance for speech and language functions. It could be argued that the faster processing of facial stimuli

---

[1] This research was supported by an award from the Australian Research Grants Committee.

[2] Requests for reprints should be sent to Gina Geffen, Department of Psychology, Monash University, Clayton, Victoria 3168, Australia.

directed to the right hemisphere is dependent on the matching aspect of the task (Geffen et al., 1971), whereas identification produces left hemisphere superiority. A same–different RT task which investigates both verbal and visual (spatial) encoding with similar material and of comparable complexity is required to elucidate the basis of the right hemisphere superiority for matching faces.

Laterality effects might also be accounted for by attentional bias rather than perceptual asymmetry (Oxbury, Oxbury, & Gardner, 1967). Myers (1970) draws a distinction between voluntary and involuntary control of the direction of attention. If right-handed *S*s attend voluntarily to the right ear or RVF in the absence of instructions to behave differently, then knowledge of which visual field (or ear) would be stimulated should reduce any laterality effects. An alternative view is that attention is involuntarily directed toward the visual field, or ear, contralateral to the hemisphere dominant for the class of stimuli presented. This would be the RVF or right ear for verbal stimuli and the LVF and left ear for visuospatial and musical stimuli. Processes that might underlie differential expectancies and their effects are discussed by Kinsbourne (1970). This model would predict increased laterality effects when *S* either knows which side will be stimulated, or whether a name or a physical match is required, or both. The present experiment tested the effects of expectancy on lateral asymmetry in physical and name matching.

METHOD

Procedure.—The letters used were A, E, M, and R printed in the upper- and the lowercases. They were chosen because of their low acoustic confusability, and because it was felt that any pair could be visually matched with equal facility. Unlike previous experiments (Posner & Mitchell, 1967), the letters were deliberately chosen so that the lowercase letters were all of the same size, without tails, and were also drawn the same size as the uppercase letters. It was considered important that matching should not be facilitated by a possible size cue. The letters were hand-drawn each to a height of 12 mm., width 10 mm., and stroke thickness 2 mm. Photographic reproductions of two adjacent letters were pasted on cards. The angular separation between the fixation point and the center point between the two letters was 2.75° and each stimulus pair subtended a visual angle of 1° at a viewing distance of 130.8 cm.

The *S*'s task was to judge whether a pair of letters, tachistoscopically presented, was the same or different. A pair of letters could be judged "same" if they were physically identical (AA, aa) or if they were the same in name only (Aa).

Items were presented in a three-channel Scientific Prototype tachistoscope, Model GB. Luminance was matched on two fields and checked by a Salford Electrical Instruments spot photometer. It was kept constant at 137.0 cd/m². A fixation point in the center of the preexposure field was present at all times except during presentation of a stimulus pair. A warning tone (1 kHz. and 30 db.) sounded for 500 msec., and was immediately followed by a stimulus pair which was presented for 150 msec. either to the left or the right of fixation. Onset of the stimulus pair triggered an electronic timer which was stopped when *S* pressed buttons on the response panel in front of him. Both hands were used since a bilateral response made it more likely that both hemispheres were in control and response asymmetry would therefore be minimized. The forefingers were placed on the two buttons closer to *S* and the third digits on the two buttons further away. The *S* depressed the closer buttons (forefingers) if he thought the two letters were the same and the other two buttons (third digits) if he thought the two letters were different. Depression of the first key stopped the timer, but *S*s were informed that a simultaneous response was necessary. Whether the left or right digit led was not monitored. All *S*s held their hands palm down with elbows on a table and appropriate fingers resting on top of the buttons.

There were three conditions. In the first, SAME was defined as a pair of letters that were physically identical (e.g. AA), and *S*s were told that letters which were similar in name only (Aa) would never be presented. In this series of purely physical matches, there were 96 trials, half of which required a SAME response, and the other half a DIFFERENT response. Half of the stimulus pairs occurred in the LVF and half in the RVF. The SAME and DIFFERENT items and their side of occurrence were pseudorandomized, with the constraint that any repeated pairs which could produce serial repetition effects occurred as often on the left as on the right. The *S*s had been informed that SAME and DIFFERENT stimuli occurred equally often, that stimuli were presented as often on the left as on the right, and that the order of presentation was random.

The second condition was similar in all respects to the first except that SAME was defined as a pair of letters having the same name (Aa), and *S*s were told that letter pairs which were physically identical would never be presented. The third condition was a mixture of the first and second. The same number of trials were presented, but half of the SAME stimulus pairs were physically identical (as well as necessarily having the same name, e.g., AA), while the rest of the SAME trials were similar only in name (Aa).

The three conditions described above were also presented a second time. However, instead of randomizing the side of presentation of stimulus pairs, 48 pairs were presented on the left, then 48 on the right, i.e., in blocks. Half of the *S*s received the blocked presentation first, and the other half received the randomized side of presentation first. Half of the *S*s started the blocked presentation on the right; thereafter left and right sides of presentation were alternated in blocks of 48 trials. The third, mixed condition, whether it was randomized or blocked with respect to side of presentation, always followed the other two conditions, physical and name matching. If it had preceded the pure conditions, *S*'s expectancies for either name matching or physical matching might have been prejudiced. Half of the *S*s received pure physical matching first and the other half pure name matching first. Sixteen practice trials preceded each condition.

A TV camera (Ikegami Tsushinki Co.) fitted with a 100-mm. lens together with an extra meniscus lens was placed in the third field of the tachistoscope. A front-surface mirror, located in front of the lens, and out of *S*'s line of fixation to the other two stimulus fields of the tachistoscope, reflected the image of *S*'s left eye through 90° into the camera lens. The *S*'s eye was illuminated by a tungsten lamp, whose emitted light was cooled by a water cell (3.5 mm. thick) and passed through a translucent white diffuser (12.7 × 10.2 cm). This produced a diffuse illumination of 683.2 cd/m^2. The lamp was not illuminated continuously, but merely for the period immediately preceding the warning tone until after *S*'s response. The *S*s maintained a constant head position by leaning their foreheads against a headrest attached to the tachistoscope and placing their chins on a chin rest. The boundary of the image of *S*'s iris was marked on the screen of the TV monitor (Astor, Model VM8B) with a felt pen while he fixated the central fixation point. The TV monitor magnified the iris threefold. A 3-mm. deflection on the screen was reliably detected. The *E* always restored *S*'s position after any rests or pauses.

If *S* made an error, or his RT exceeded an arbitrarily set limit of 1 sec., or if a lateral eye movement was observed on the TV monitor during a trial, then these trials were discarded. If discarded items exceeded 16% of the total number of trials per condition, they were repeated at the end of that condition. The repetitions were given in a sequence with the same constraints as the original, and *S*s were not aware that they were repeating errors. At the start and finish of the experiment, *S*s were given eight simple RT or calibration trials to assess the magnitude of any differences in response times between the second and third digits. These trials consisted of stimulus pairs presented to the left or right of fixation; on half of the trials the letter pairs were physically identical, the other half being different. The *S*s were told prior to the presentation of each stimulus pair whether a SAME or DIFFERENT response was required.

The *S*s were tested individually by two *E*s, one of whom changed stimulus cards while the other noted RTs and started the cycle for each trial. The second *E* also monitored the TV screen. The entire experiment took 4 hr. to perform, each *S* being tested in hourly sessions, run usually on 4 successive days.

Subjects.—Twelve right-handed *S*s with no known close left-handed relatives were employed. There were 6 males and 6 females. They were paid undergraduate volunteers who were without uncorrected visual defects. Each *S* was first given an acuity test using a Bausch and Lomb orthorater for near and far vision, with left and right eyes tested separately. None of the 12 *S*s had more than two points difference between eyes for both near and far vision. This pretest was used to control for any possible contaminating effects of eye dominance on laterality differences.

RESULTS

A *t* test for correlated samples showed that the differences between mean RTs for the second and third digits during the calibration trials was not significant, $t(11) = 1.5, p > .05$. The mean RTs summed across *S*s were 307.9 msec. for the second digit and 295.5 msec. for the third digit. This is in contrast to the finding of Bradshaw and Perriment (1970) that the second digit was faster than the third. However, their *S*s were responding unilaterally and with their fingers in a left–right orientation. The up–down positions of the fingers in the present study probably allows greater leverage for the third digit, thus reducing the difference in RT between them.

Table 1 shows mean RTs summed across *S*s and blocked and randomized side of presentation, which were shown not to differ. A five-way within-*S*s analysis of variance was carried out on the data for the pure physical- and pure name-matching conditions. A similar analysis was performed on the SAME responses only, in both the two pure and mixed conditions. Finally, a four-way within-*S*s analysis of variance was done on the different responses in the pure and mixed conditions.

Visual field differences.—It was predicted that name matching would be faster when these stimuli were presented in RVF, while letters to be physically matched would be

TABLE 1

Mean Reaction Times for 12 Subjects Summed Across Blocked and Randomized Side of Presentations Showing Visual Field Differences

Cond.	Type of matching	LVF	RVF	L–R
Pure	Physical			
	Same	535	547	−12
	Different	601	613	−12
	Name			
	Same	680	657	23
	Different	717	687	30
Mixed	Physical			
	Same	509	540	−31
	Name			
	Same	641	626	15
	Different	644	656	−12

Note.—LVF = left visual field; RVF = right visual field; L–R = LVF minus RVF.

responded to more quickly when projected in LVF. Hence the Visual Field × Physical or Name Matching interaction must be examined. In all three analyses, this interaction was found to be highly significant; for the SAME responses at F (1, 11) = 12.25, $p < .01$. For the pure conditions (SAME and DIFFERENT responses), the Visual Field × Physical or Name Matching interaction was significant at F (1, 11) = 11.83, $p < .01$; the similar interaction for different responses was significant at F (2, 22) = 11.18, $p < .001$. The mean RTs for each *S* for the left and right visual field stimuli for physical and name matching were pooled across pure and mixed and blocked and random conditions and examined. Ten of the 12 *S*s responded more quickly to LVF stimuli in the physical matching conditions; conversely 10 of 12 *S*s have faster mean RTs for name matches in RVF.

Different responses in the mixed conditions were examined separately, since the direction of the visual field difference for these RTs was not predicted. These items were processed more quickly when presented in LVF, 9 of the 12 *S*s showing a difference in this direction, while the other 3 had faster RTs for the RVF stimuli. The advantages of the RVF for name matching and the LVF for physical matching are stable. There is a lack of significant interaction between visual field and SAME or DIFFERENT response, $F < 1.0$; and between visual field and blocked or randomized side of presentation, F (1, 11) = 1.98, $p > .05$, for analysis of SAME responses; and $F < 1.0$ in the analysis of DIFFERENT responses. The LVF superiority for physical matching of SAME stimuli (AA) is greater in the mixed than in the pure condition (31-msec. difference compared to 12-msec. difference), and this Visual Field × Pure or Mixed Condition interaction was significant, F (1, 11) = 5.15, $p < .05$.

Errors were analyzed in the same way as RTs. There was no significant difference between the left and right visual field in the error rate; 6.5% and 6.9%, respectively, $F < 1.0$ in all three analyses.

Physical and name matching.—Physical matching was significantly faster than name matching: in the pure conditions, F (1, 11) = 126.44, $p < .001$. Letter pairs which were physically identical were matched 127.6 msec. faster than those which were similar in name only. Different items were likewise processed more quickly (by 94.5 msec.) in the pure physical-matching condition compared to the pure name-matching condition. For SAME responses in the mixed condition, the physical matching was 109.2 msec. faster than name matching, F (1, 11) = 18.81, $p < .01$. The DIFFERENT RTs in the mixed condition (mean RT = 650 msec.) were faster than those in the pure name condition (mean RT = 702 msec.) but slower than the RTs to DIFFERENT items in the pure physical-matching condition (mean RT = 607 msec.). This main effect was significant, F (2, 22) = 49.69, $p < .001$.

There were significantly more errors and long RTs during name-matching than in the physical-matching conditions. This was true for all 12 *S*s, and all three analyses. For the pure conditions, F (1, 11) = 26.95, $p < .001$; for the SAME responses (pure and mixed), F (1,11) = 32.36, $p < .001$; and the DIFFERENT responses, F (2, 22) = 20.34, $p < .001$. For SAME items (pure + mixed) which were name matches, the error rate was 9.9%, compared to 3.5%

for physical matches. Similarly, the error rate for different items was 5.3% during name matching, 1.8% during physical matching, and 2.7% in the mixed condition.

SAME and DIFFERENT RT.—The SAME responses were significantly faster than different responses by 50 msec. (pure physical + pure name matching), $F\,(1, 11) = 27.72$, $p < .001$. The Type of Matching × Type of Response interaction is significant, $F\,(1, 11) = 21.51$, $p < .001$. This might be due to the smaller difference between SAME and DIFFERENT RTs for name than physical matching (33 msec. compared to 66 msec.).

There was no significant difference in error rate between SAME and DIFFERENT responses (6.7% and 6.5%, respectively), $F\,(1, 11) = 2.07$, $p > .05$. This may represent a lack of strong response biases.

Pure and mixed conditions.—The sequences which contained both physical and name matches were presented after pure physical and pure name matching. As such, the mixed stimuli received faster RTs, $F\,(1, 11) = 15.36$, $p < .01$. Responses to physically identical letter pairs in the mixed condition were, on the average, 16.2 msec. faster than those in the pure condition, while matching based on name similarity was an average of 34.6 msec. faster in the mixed condition than in the pure name condition. These differences are consistent across *S*s, all of whom responded more quickly to same items in the mixed condition.

The error rate was slightly, but significantly lower in the mixed than in the pure conditions for SAME responses (7.9% and 5.5%, respectively), $F = 6.54$, $p < .05$. Posner and Mitchell (1967) have shown a general increase in RTs with mixed lists and explained this in terms of the increase in overall difficulty of the task with mixed lists. The findings here may indicate that the difficulty of the task decreases with practice.

Random and blocked side of presentation.—The main effects of type of presentation were not significant, $F < 1.0$ for both the pure conditions and SAME RTs (pure + mixed) analyses, and $F\,(1, 11) = 3.49$, $p > .05$ for the DIFFERENT RTs analysis. There was a significant Type of Matching (pure, name, or mixed) × Type of Presentation interaction for the DIFFERENT items, $F\,(2, 22) = 4.33$, $p < .05$. This was due largely to the faster RTs to DIFFERENT items during random rather than blocked presentation in the mixed condition. The mean RTs summed across *S*s were 624 msec. and 678 msec., respectively. Nine of the 12 *S*s showed this pattern of response. During trials blocked for side of presentation, it was stressed to *S*s that they should maintain central fixation especially during the warning interval. This requirement might have caused the paradoxical increase in RTs. None of the interactions with blocked or randomized side of presentation were significant. There was no significant difference in error rate between the two types of presentation (7.3% for blocked and 6.0% for random); $F < 1.0$ for both the pure conditions, and SAME responses (pure + mixed); and $F\,(1, 11) = 3.63$, $p > .05$ for the different responses.

Eye movements.—The occurrence of eye movements was rare, .5% in the total experiment. They were evenly distributed between the visual fields, 18 occurred prior to LVF stimuli and 19 before RVF stimuli. There were 21 eye movements during blocked presentation and 16 during randomized side of presentation. The overall occurrence was too infrequent to reveal statistically significant effects. The rarity of eye movements indicates that *S*s do fixate centrally when this is requested.

DISCUSSION

Perceptual asymmetry between the hemispheres has been confirmed. Letter pairs which were similar in name were processed faster when directed to the left, language, hemisphere. The mean visual field differences in RT (about 20 msec.) were very similar to previous findings (Geffen et al., 1971). Conversely, physically identical letter pairs were processed more rapidly when presented to the LVF and thus to the right hemisphere. On the other hand, Gazzaniga (1970), in a preliminary report, found faster name matching with stimuli in the RVF but no visual field difference with stimuli which were to be matched physically.

These visual field differences may be interpreted as evidence for hemispheric specialization; the left, language hemisphere processing name matches, and the right, spatial hemisphere processing items which are physically identical. The increase in RT when name matches were presented in the LVF and physical matches in the RVF may be the result of transcallosal transmission of information, or it may reflect less efficient, and therefore slower, processing of verbal material by the right hemisphere and visuospatial material by the left hemisphere. It has recently been argued that both transcallosal transmission times and less efficient processing of inappropriate stimuli are implicated (Geffen et al., 1971).

White (1971) has challenged the cerebral dominance explanations of lateral differences in visual perception. He found that recognition of single English capitals was better when these were presented in the RVF, but he also found RVF superiority for recognizing the orientation of a single line. White suggests that "for certain classes of stimuli, visual laterality differences may be subserved by a selective contour-tuning mechanism [p. 207]." However, these findings may well result from the use of a vocal response by White's *S*s both in reporting which letter was displayed and in calling out a numeral corresponding to a given line orientation. Filbey and Gazzaniga (1969) have shown that a vocal response can produce faster RTs in the detection of the presence or absence of a dot when this is presented in the RVF. A manual response in the dot-detection task yielded no visual field differences. Although White (1971) did not measure RT, it may well be that when a signal is directed to the right hemisphere it has to be transferred to the left hemisphere for initiation of the vocal response, and there is information loss or signal degradation during the transfer across the commissures.

The finding of faster processing of physically identical stimuli when these are directed to the right hemisphere suggests that the type of task is more important than the type of stimulus in determining hemispheric asymmetry. When analysis of visual patterns is required, the right hemisphere is better.

The complete lack of significant interactions between blocked or random side of presentation and visual field indicates that *S*'s knowledge of which side will receive the stimulus neither increases nor decreases lateral asymmetry. This means that attending to the left or right, with visual presentation, does not tune in the contralateral hemisphere (cf. Kinsbourne, 1970). The second manner of manipulating *S*'s expectancy involved either presenting pure name matching or physical matching, or using mixed presentation where *S* could not predict whether a match based on physical or name identity would be required. Only in one instance did this manipulation have a significant effect on visual field differences in RT. In the mixed condition, the LVF superiority for matching on the basis of physical identity was more marked than in the pure condition. The increased stimulus uncertainty in the mixed condition may produce more apparent sharing of the processing load by the two hemispheres, each specializing in different functions. Thus there is no evidence for increased attention to the opposite side when a cognitive set proper to one hemisphere is adopted. On the contrary, lateral asymmetry is increased when either a physical or a name match is expected.

As in previous studies (e.g., Posner & Mitchell, 1967), name matching took significantly longer than matching on the basis of physical identity. However, SAME responses are usually faster than DIFFERENT responses only when the letters to be matched are presented successively, not simultaneously (e.g., Forrin & Morin, 1969; Nickerson, 1965, 1968). Posner and Mitchell (1967) have suggested that differences in RT between SAME and DIFFERENT responses may be dependent on the particular population of materials used. It is likely that the use of uppercase and lowercase letters of the same size in the present experiment would increase the time taken to respond to a pair of different stimuli, merely because a size cue has been eliminated.

The results of this experiment suggest that spatial comparisons of letters favor processing in the right hemisphere, whereas the left hemisphere is superior when a comparison involves verbal labels. The differences in response times are suggestive of interhemispheric transfer of information; however, less efficient processing of verbal material by the right hemisphere and spatial material by the left hemisphere could also account for these differences.

REFERENCES

Bradshaw, J. L., & Perriment, A. D. Laterality effects and choice reaction time in a unimanual two-finger task. *Perception & Psychophysics*, 1970, 7, 185–188.

EGETH, H. Laterality effects in perceptual matching. *Perception & Psychophysics*, 1971, **9**, 375–376.

FILBEY, R. A., & GAZZANIGA, M. S. Splitting the normal brain with reaction time, *Psychonomic Science*, 1969, **17**, 335–336.

FORRIN, B., & MORIN, R. E. Recognition times for items in short- and long-term memory. *Acta Psychologica*, 1969, **30**, 126–141.

GAZZANIGA, M. S. *The bisected brain*. New York: Appleton-Century-Crofts, 1970.

GEFFEN, G., BRADSHAW, J. L., & WALLACE, G. Interhemispheric effects on reaction time to verbal and nonverbal visual stimuli. *Journal of Experimental Psychology*, 1971, **87**, 415–422.

KINSBOURNE, M. The cerebral basis of lateral asymmetries in attention. *Acta Psychologica*, 1970, **33**, 193–201.

KLATZKY, R. L. Interhemispheric transfer of test stimulus representations in memory scanning. *Psychonomic Science*, 1970, **21**, 201–203.

KLATZKY, R. L., & ATKINSON, R. C. Specialization of the cerebral hemispheres in scanning for information in short-term memory. *Perception & Psychophysics*, 1971, **10**, 335–338.

OXBURY, S. OXBURY, J., & GARDINER, J. Laterality effects in dichotic listening. *Nature*, 1967, **214**, 742.

MYERS, T. F. Asymmetry and attention in phonic decoding. *Acta Psychologica*, 1970, **33**, 158–177.

NICKERSON, R. S. Response times for "same"–"different" judgments. *Perceptual and Motor Skills*, 1965, **20**, 15–18.

NICKERSON, R. S. Note on "same"–"different" reponse times, *Perceptual and Motor Skills*, 1968, **27**, 565–566.

POSNER, M. I., & MITCHELL, R. F. Chronometric analysis of classification. *Psychological Review*, 1967, **741**, 392–409.

WHITE, M. J. Visual hemifield differences in the perception of letters and contour orientation. *Canadian Journal of Psychology*, 1971, **25**, 207–212.

(Received October 1, 1971)

urnal of Experimental Psychology
972, Vol. 95, No. 1, 32–38

SENSORY AND ACTIVE STORAGE OF COMPOUND VISUAL AND AUDITORY STIMULI[1]

NEAL E. A. KROLL,[2] STANLEY R. PARKINSON,[3] AND THEODORE E. PARKS

University of California

While *S*s shadowed female-voiced letters, five memory letters were presented either auditorily (in a male voice), visually, or both auditorily and visually. A simple additive formula was used to predict recall performance following concurrent visual and auditory presentation. The formula accurately predicted performance for immediate recall, indicating independent visual and auditory storage systems for very brief time periods. However, following retention intervals of 2, 6, and 20 sec., the predicted scores were higher than those actually observed. Deviations from the formula at the longer retention intervals were explained by assuming (*a*) the existence of a limited duration precategorical acoustic storage (PAS) system, and (*b*) that rehearsals were primarily visual.

Atkinson and Shiffrin (1968, 1971) have proposed a model of memory which provides both an excellent organization of existing data and also an extremely useful framework for guiding current research. According to that model, recall performance over brief retention intervals is often a joint function of several memory systems. First, it appears likely that memory can be divided vertically into (*a*) a very brief store which requires little effort on the part of *S*, (*b*) an active store which consists of frequent rehearsals of the information being stored, and (*c*) a long-term store in which information is passively maintained but highly susceptible to proactive interference from memories of earlier presented information of a similar nature. It also appears that memory can be divided horizontally on the basis of the nature of the memory code, i.e., the sensory modality, if any, to which it is most analogous.

One important implication of this model concerns the interval over which information will be stored in a form similar to the way in which that information was received. Generally such storage will occur for at least a very brief period following stimulation (i.e., during storage in the very brief store). However, storage beyond these few seconds usually must be actively maintained through the process of rehearsal. Depending on the existing conditions, *S* may choose not to maintain an active store in a given modality even though input had occurred in that modality. This point has already been well established with respect to the adult human's propensity to store visually presented stimuli via a visual memory code for only a very brief period and to rely on an auditory–verbal-linguistic (AVL) trace if longer retention is required (e.g., Averbach & Sperling, 1961). Since recent studies make it clear that *S*s can maintain an active visual memory (e.g., Kroll, Parks, Parkinson, Bieber, & Johnson, 1970; Parks, Kroll, Salzberg, & Parkinson, 1972), the prevalance of the AVL form of active storage in earlier studies apparently reflects an *S* preference based on the relative ease of AVL rehearsal under many experimental conditions.

The central point of the present research concerns the fate of each of the two components of a memory stimulus that is presented both visually and auditorily concurrently under conditions where active AVL storage is relatively difficult. That is, *S*s were required to shadow (repeat aloud) a rapid series of auditory stimuli during the retention interval. Under these conditions, it would be expected that brief sensory store would occur for each component of the compound stimulus, the visual component in "iconic" store (Neisser, 1967, p. 20) and the auditory component in "Precategorical Acoustic Storage" (PAS) (Crowder & Morton, 1969). If so, then recall performance immediately following such presentations ought to be a joint function of both the auditory and the visual traces of the respective components of the memory stimulus. Specifically, the probability of recalling a compound stimulus, $P(V + A)$, ought to be predictable from the simple additive formula: $P(V + A) = P(V) + P(A) - P(V)P(A)$; where $P(V)$ and $P(A)$ are the probabilities, estimated on spearate trials, that *S* can recall a stimulus he has only seen and a stimulus he has only heard, respectively. Each such single-modality stimulus would supposedly have the same probability of entering and remaining in the appropriate very brief store as would the corresponding component of a compound stimulus.

---

[1] The research reported in this paper was supported by Grants MH 19327 and MH 49532 from the National Institute of Mental Health.

[2] Requests for reprints should be sent to Neal Kroll, Department of Psychology, University of California, Davis, California 95616.

[3] Now at Arizona State University.

However, when the retention interval is increased beyond a few seconds, *S* can no longer rely on iconic and PAS traces but must rely, instead, on active memory. At this point, the critical question concerns the hypothesized presence of alternative modes of maintaining an active trace. If, in fact, parallel and independent rehearsal mechanisms are available, and if shadowing makes AVL rehearsal difficult or relatively ineffectual, *S* should choose to actively maintain only the visual component of a compound memory stimulus. Thus, once PAS is lost, retention of compound stimuli, $P(V + A)$, ought to approximate the retention of what were solely visual stimuli, $P(V)$.

METHOD

Design.—On all experimental trials the shadow letters were presented in a female voice at a two/sec rate binaurally through headphones. Each trial contained five memory letters and each of the four experimental conditions differentiated memory letters from shadow letters differently.

In the Auditory condition, five female-voiced shadow letters were removed and replaced by a male voice saying the memory letters at a two/sec rate. The *Ss* were told to shadow both the female- and the male-voiced letters but to remember only the male-voiced letters.

In the Visual condition, five visual letters (8.5 cm. in height) were presented individually, left to right, on five readout units located approximately 1.8 m. from *S* and approximately at eye level. These visual letters were presented at a two/sec rate, each with an on-time of approximately 200 msec., while the female voice said different shadow letters. The *Ss* were instructed to continue shadowing the female voice while they watched and remembered the visually presented letters.

The Visual + Auditory condition employed the tape recordings of the Auditory condition and, in addition, presented the memory letters visually as the male voice said them. Thus in this condition *S* saw, heard, and said the memory letters.

The Visual Read condition employed the tape recordings of the Visual condition but was preceded by a different set of instructions. Here *Ss* were instructed to stop shadowing the female voice during memory-letter presentation and to read the visually presented memory letters instead. Thus *Ss* saw and said the memory letters while hearing irrelevant female-voiced letters. This condition was included as a rough check on attentional factors. That is, the Visual + Auditory condition could result in better memory than the Visual condition for either one of two reasons: an additional memory system might be brought into play (as hypothesized), or *S* might pay more attention to the memory letters. That is, in the Visual condition, *Ss* must attend to both the visual memory letters and to the shadowing task, while in the Visual + Auditory condition the shadowing task is concerned with the same letters as the memory task. The Visual Read condition allows *S* to devote full attention to the memory letters, but does not present the memory letters auditorily. Since *Ss* might possess an auditory memory of their own voice, different results might be expected had *Ss* remained silent during visual memory-letter presentation. However, the present condition was used as it was believed to be the stronger control; i.e., if the Visual + Auditory condition was superior to the Visual Read condition (as hypothesized), it should also be superior to a visual, silent read condition (Murray, 1966).

Three groups of 12 *Ss* each were run. One group had 20 trials each for the Visual, Auditory, and Visual + Auditory conditions; 10 trials with a 0-sec. retention interval and 10 trials with a 20-sec. retention interval. Another group had the same three conditions but with each having 10 trials with a 2-sec. retention interval and 10 trials with a 6-sec. retention interval. The third group had 20 trials each of the Visual, Auditory, and Visual Read conditions; 10 trials with a 0-sec. and 10 trials with a 20-sec. retention interval.

The conditions were blocked with each condition being preceded with its own short set of instructions. Within each block of 20 trials, the 10 trials at the shorter interval and the 10 trials at the longer interval were interspersed in such a way as to prevent *Ss* from knowing in advance the length of the retention interval. The first 4 trials (2 of the shorter interval and 2 of the longer interval) of each of the conditions were considered practice trials and were not scored.

All 36 *Ss* received the same 60 shadow lists and memory letters in the same order, but between *Ss* the assignment of the 20-trial blocks to conditions and the assignment of trials within blocks to retention intervals was such that, over all *Ss*, each block was run under each condition and each trial was run at each retention interval.

Materials.—Each readout unit was capable of displaying 1 of 12 letters. Memory-letter sets of 5 were chosen from these possibilities such that (*a*) no two letters of a set were the same, (*b*) each letter in each position was used with approximately equal frequency, (*c*) adjacent trials had sets as dissimilar as possible, (*d*) the probability of an association was low, and (*e*) acoustic confusions within the set were held to a minimum.

Letters used in the shadow list for a given trial could include any letters except "W" and those letters used as memory letters on that trial. Each shadow list contained 16 lead-in shadow letters, the appropriate memory-letter presentation condition, and the appropraite number of letters to result in the required retention interval. Shadow lists were played from one channel of a tape recorder while tones on the other channel synchronized the visual presentations via a sensitive relay.

Procedure.—Each *S* served in a single 1½-hr. session, but was allowed short rest periods between the various portions of the session. The session began with approximately 30 min. of shadow practice similar to that given *Ss* in earlier experiments (e.g., Kroll et al., 1970). Just prior to his first experimental condition, *S* was instructed to attempt to write the five memory letters, in order, on his answer sheet at the end of each trial. After writing his answer, he was to advance the answer sheet under the cover on the arm of his chair such that his answer would be covered for succeeding trials. A 15-sec. intertrial interval was provided for the recall response.

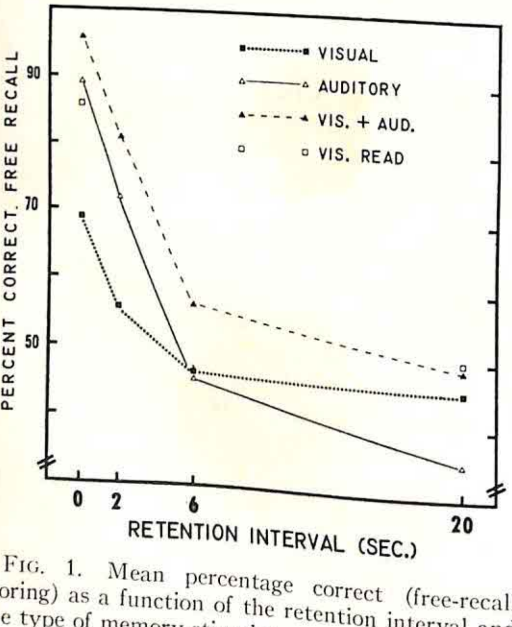


FIG. 1. Mean percentage correct (free-recall scoring) as a function of the retention interval and the type of memory stimulus presentation.

RESULTS AND DISCUSSION

Although *Ss* had been instructed to attempt recalling the memory letters in serial order, the following results are based on a free-recall scoring of the data. Results based on serial recall scoring had similar patterns and only differed from those based on free-recall scoring by having (*a*) lower overall scores, (*b*) more pronounced serial position curves, and (*c*) somewhat reduced magnitudes of effect for variables other than serial position.

Figure 1 presents the percent correct letters for each of the conditions of the experiment. In this figure, the results for the Visual and for the Auditory conditions at 0 and 20 sec. are averaged over 24 *Ss*, 12 of whom had Visual + Auditory as their third condition and 12 of whom had Visual Read as their third condition. All other points on this figure represent 12 *Ss*.

Recall performance following visual and auditory presentation of the memory letters were compared over all 24 *Ss* having retention intervals of 0 and 20 sec. At 0 sec., the Auditory memory letters were recalled better than Visual letters, F (1, 23) $= 48.3, p < .01$; while at 20 sec. the Visual memory letters were better recalled, F (1, 23) $= 10.9, p < .01$. These results replicate those of Parkinson (1972) and those of Peterson and Johnson (1971) in finding that recall following an auditory presentation of multiple memory letters is superior with short retention intervals while recall following a visual presentation is superior with longer intervals when the intervals are filled with a verbal task.

The finding of an early recall superiority following auditory presentation is in agreement with the Crowder and Morton (1969) hypothesis that PAS is longer lasting than iconic storage. The finding of a later recall superiority following visual presentation suggests that *Ss* find visual rehearsals more efficient than AVL rehearsals during a concurrent verbal task. This, in turn, raises the question of major concern to this ex-

TABLE 1

PERCENTAGE OF CORRECT RESPONSES OBSERVED UNDER VISUAL, AUDITORY, AND VISUAL + AUDITORY CONDITIONS

Cond.	0-sec.	2-sec.	6-sec.	20-sec.
Observed				
(V)	69	56	47	47
(A)	91	72	46	35
(V + A)	96	81	57	49
Predicted[a]				
$(V + A)'$	97	88	71	66
$(V + A)' - (V + A)$	1	7	14	17
$t(11)$	.99	3.63**	3.47**	5.09***
$(A) - (V + A)$	−5	−9	−11	−14
$t(11)$	2.66*	3.25**	3.98**	4.41**
$(V) - (V + A)$	−22	−25	−10	−2
$t(11)$	5.85***	6.75***	2.34*	.65

Note.—(V) = Visual, (A) = Auditory, (V + A) = observed Visual + Auditory, (V + A)′ = predicted Visual + Auditory.
[a] $P(V + A)' = P(V) + P(A) - P(V)P(A)$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

periment: If the memory stimulus is presented to both modalities, will *S* retain both memory traces or, if he retains only one, which will he retain? The model advanced in the introduction suggested the following answers: (*a*) Since immediate recall depends largely on memory traces which are independent of *S*'s rehearsal strategy, immediate recall should fit the simple additive formula, $P(V + A) = P(V) + P(A) - P(V)P(A)$. (*b*) Since delayed recall will depend increasingly on *S*'s rehearsal strategy as the retention interval is increased, and since it is assumed that *S* will, while shadowing, rehearse visually, delayed recall under the Visual + Auditory condition should become less like the prediction of the simple additive formula and more like the recall performance under the Visual condition.

Table 1 presents the free-recall performance (percent correct recall) for the 12 *S*s tested at 0- and 20-sec. retention intervals and for the 12 *S*s tested at 2- and 6-sec. intervals; all of these *S*s having the Visual + Auditory condition. The observed Visual + Auditory performances, $P(V + A)$, are in Row 3, while the scores predicted by the simple additive formula, $P(V + A)'$, are in Row 4. Observed Visual + Auditory scores were subtracted from predicted Visual + Auditory scores (Row 5), from the Auditory scores (Row 7), and the Visual scores (Row 9). The *t* tests testing these differences are in Rows 6, 8, and 10, respectively.

As predicted by *a* above, immediate recall performance under the Visual + Auditory condition is significantly better than under either the Visual or the Auditory conditions, but is virtually identical to the performance predicted by the simple additive formula. As predicted by *b* above, increasing the retention interval results in the performance under the Visual + Auditory condition becoming less like the prediction of the simple additive formula and more like the performance under the Visual condition. While the prediction of no difference between $P(V + A)$ and $P(V + A)'$ at 0 sec. and between $P(V + A)$ and $P(V)$ at 20 sec. raises the awkward problem of testing the model via accepting the null hypothesis, it is clear that (*a*) these differences *are* nearly zero, and (*b*) the experiment is quite sensitive, as evidenced by the finding of significance for all differences in Rows 5, 7, and 9 except those two differences predicted to be zero.

The proposed model attributes the initial high recall under the Visual + Auditory condition to an unrehearsed auditory

(PAS) component. That it was due to an auditory component, rather than increased attention or an articulatory component, is supported by the significant difference between the Visual + Auditory condition and the Visual Read condition at 0 sec., $F\ (1, 22) = 5.73, p < .05$. (However, it must be admitted that the Visual Read condition is better than the Visual condition, $t\ (11) = 4.41, p < .01$. This difference may be due to increased attention, articulation, or to PAS resulting from *S*'s own voice.)

If the proposed model is correct, it should be possible to solve for P(A) in the additive formula, using the observed values of P(V + A) and P(V), to obtain the *effective* auditory component in the recall of stimuli presented both visually and auditorily. The plot of this effective auditory component over time is given in Fig. 2 for both free-recall and serial recall scoring to demonstrate the agreement between the two scoring procedures.

The computed auditory memory component depicted in Fig. 2 is interpreted as evidence for a PAS memory component which lasts approximately 8 sec. in the presence of shadowing as estimated from the intersection of a line through the 0-, 2-, and 6-sec. free-recall points with a horizontal line through the 20-sec. point. Figure 2 is based on performance averaged over all five positions of the memory stimulus. Similar calculations for each individual position resulted in estimates ranging from 7 to 8½ sec. While these estimates are somewhat longer than the under 5.3-sec. estimation of Glucksberg and Cowen (1970), the correspondence seems acceptable given the great differences between the two experiments. For example, Glucksberg and Cowen attempted to prevent *S*s from attending to the memory stimulus and, on the basis of subjective reports, were fairly successful in doing so. In the present study, *S*s reported listening to the memory letters (i.e., they reported shadowing the male voice and not merely reading the visual letters off the screen), and were therefore likely to be giving a good deal of attention to the auditory stimuli. This difference in degree of attention paid to the auditory memory stimulus may explain the slight difference in the duration estimates between the two studies (5 sec. vs. 8 sec.) and the large differences in the estimates of initial strength (26% vs. 87%).

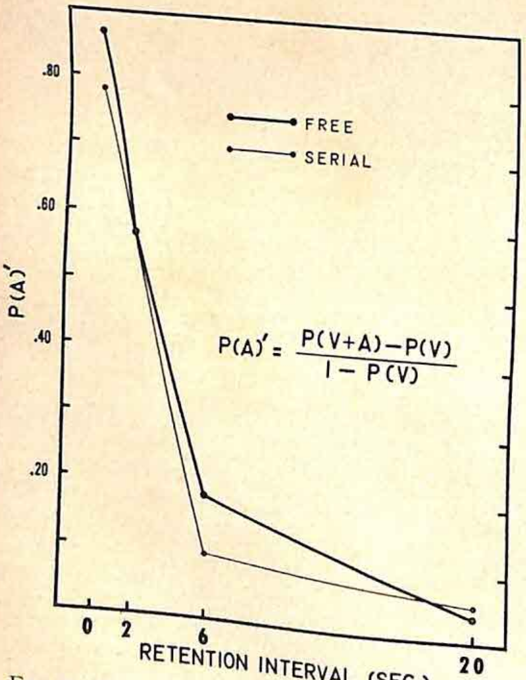


FIG. 2. Hypothesized effective auditory memory within the Visual + Auditory condition as a function of retention interval for both free-recall and serial recall scoring.

The serial position curves for Auditory, Visual, and Visual + Auditory conditions—at 0-, 2-, 6-, and 20-sec. retention intervals—are presented in Fig. 3. (As in Fig. 1, the Auditory and Visual conditions at 0- and 20-sec. retention intervals represent 24 *S*s while all other conditions represent 12 *S*s.) These curves also seem to support the model. At 0 sec., the Visual + Auditory serial position curve resembles a rough combination of the Auditory and Visual 0-sec. curves. At 20 sec., the Visual + Auditory serial position curve is quite definitely more similar to the Visual curve than to the Auditory curve. This latter finding does not seem to be an artifact caused by the different overall percentage correct since the Auditory 6-sec. curve is at the same overall level and yet does not resemble the Visual 20-sec. curve nearly as much.

While the overall results of the experi-

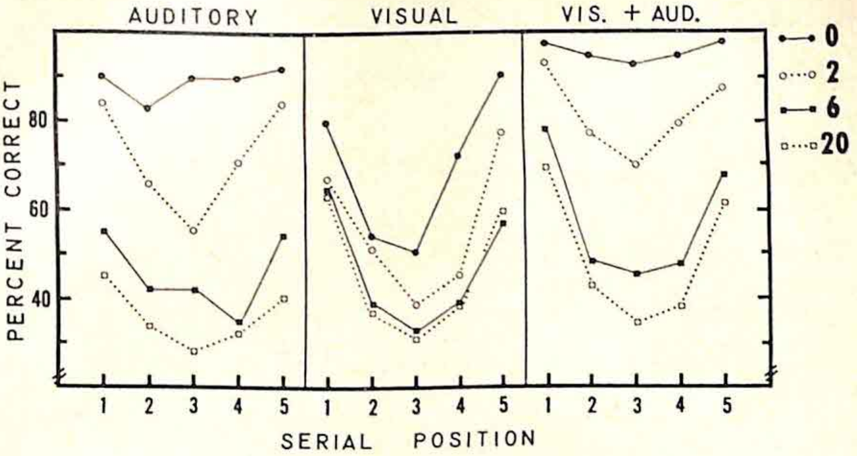


FIG. 3. Mean percentage correct (free-recall scoring) as a function of serial position (1, 2, 3, 4, or 5), type of memory stimulus presentation (Auditory, Visual, or Visual+Auditory), and retention interval (0-, 2-, 6-, or 20-sec).

ment are interpreted as supporting the PAS hypothesis, this hypothesis also predicts that the end items of a list presented auditorily will be forgotten at a faster rate than the earlier items. This prediction is typically realized when the retention intervals are filled with counting (e.g., Glanzer & Cunitz, 1966) or with counting backward (e.g., Postman & Phillips, 1965), but does not appear in the Auditory serial position curves of the present experiment. This discrepancy may be the result of additional factors introduced by the shadow task. For one thing, the shadowing probably causes much more retroactive interference for auditory memory letters than do counting tasks (e.g., Kroll & Kellicutt, 1972). This retroactive interference in turn may effect the entire list so that the early items are forgotten more rapidly than with counting tasks (cf. Glanzer & Cunitz, 1966; Postman & Phillips, 1965.) For another, attentional factors are probably quite different with the shadow task since shadowing is a concurrent task which *S* is doing before, during, and after memory-letter presentation is completed. These attentional factors may also be the cause of the extremely strong serial position effects found with visual presentation.

In summary, the results of the present experiment seem to be in very close agreement with the following hypotheses concerning memory over short retention intervals: (*a*) There exists a precategorical acoustic storage (PAS) of information which is retained several seconds without rehearsal. (*b*) At least while shadowing, *S*s typically rehearse the memory letters in a way that is related to the modality of presentation. (*c*) When the memory letters are presented both auditorily and visually, *S*s typically employ visual, rather than AVL, rehearsals while shadowing.

REFERENCES

ATKINSON, R. C., & SHIFFRIN, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*, Vol. 2. New York: Academic Press, 1968, 89–195.

ATKINSON, R. C., & SHIFFRIN, R. M. The control processes of short-term memory. *Scientific American*, 1971, **225**, 82–90.

AVERBACH, E., & SPERLING, G. Short-term storage of information in vision. In E. C. Cherry (Ed.), *Fourth London symposium on information theory*. London: Butterfield, 1961, 196–211.

CROWDER, R. G., & MORTON, J. Precategorical acoustic storage (PAS). *Perception & Psychophysics*, 1969, **5**, 365–373.

GLANZER, M., & CUNITZ, A. R. Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 351–360.

GLUCKSBERG, S., & COWEN, G. N., JR. Memory for nonattended auditory material. *Cognitive Psychology*, 1970, **1**, 149–156.

KROLL, N. E. A., & KELLICUTT, M. H. Correct recall in short-term memory as a function of covert rehearsal and of intervening task. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 196–204.

Kroll, N. E. A., Parks, T., Parkinson, S. R., Bieber, S. L., & Johnson, A. L. Short-term memory while shadowing: Recall of visually and of aurally presented letters. *Journal of Experimental Psychology*, 1970, **85**, 220–224.

Murray, D. J. Vocalization at presentation and the recall of lists varying in association value. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 448–491.

Neisser, U. *Cognitive psychology*, New York: Appleton-Century-Crofts, 1967.

Parkinson, S. R. Short-term memory while shadowing: Multiple-item recall of visually and aurally presented letters. *Journal of Experimental Psychology*, 1972, **92**, 256–265.

Parks, T. E., Kroll, N. E. A., Salzberg, P., & Parkinson, S. R. The persistence of visual memory as indicated by decision time in a matching task. *Journal of Experimental Psychology*, 1972, **93**, 437–438.

Peterson, L. R., & Johnson, S. T. Some effects of minimizing articulation on short-term retention. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 346–354.

Postman, L., & Phillips, L. W. Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, 1965, **17**, 132–138.

(Received October 20, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 39–47

RECALL IMPROVES IN SHORT-TERM MEMORY THE MORE RECALL CONTEXT RESEMBLES LEARNING CONTEXT[1]

PHILIPPE R. FALKENBERG[2]

Wake Forest University

If the contextual similarity between learning and recall within a single trial in a short-term memory (STM) paradigm is varied, recall varies proportionately. This context effect was demonstrated using variations of the Peterson-Peterson paradigm for both aurally and visually presented material, verbal and arithmetic context, and within- and between-*S*s designs. Experiments were conducted to discover whether the context effect was due to differential ability to rehearse in the recall interval, differing intertrial intervals, and different amounts of activity on *S*'s part during a trial. None of these hypotheses were supported by the data. An analysis of intrusion errors fails to support the hypothesis that the context effect is due to proactive interference. An explanation of the effect on the basis of stimulus generalization is proposed.

Changes in context influence learning and retrieval in long-term memory (LTM) in humans and animals (Bilodeau & Schlosberg, 1951; Carr, 1917; Greenspoon & Ranyard, 1957; Pan, 1926; Thomas & Jones, 1962; Watson, 1907). Do contextual changes influence learning and retrieval from short-term memory (STM), and if so what mechanisms are operating?

The Peterson and Peterson (1959) paradigm is suitable for the study of contextual influences on STM because: (*a*) the "busy" task inserted between learning and recall can be used as context; (*b*) measurements are taken within one trial on single *S*s; (*c*) the context is very different from the learning task and in no logical way can serve to assist learning or recall; on the contrary the context is deliberately chosen to hinder the rehearsal of the learned response (this is important, for all too often in studies employing context as independent variables it is not clear where the context leaves off and the to-be-learned S–R association begins); and (*d*) *S*s' attention is held to both the learning task and the contextual material.

Eight experiments are reported here: the first four experiments demonstrate the *context effect* in STM, namely that retrieval from STM improves as the context for learning is made more similar to the context for retrieval, and four more experiments examine the mechanism by which the context effect might operate.

EXPERIMENT I

In the Peterson-Peterson paradigm, *S*s learn a consonant trigram, count backward in the interval between learning and recall, then recall what they can of the trigram. The two criterion events of interest in this paradigm are the learning of the trigram and the recall of the trigram. Events that occur immediately before a criterion event can be construed as part of its context. Note that the context for the recall is very different from the context for learning, because *S*s are counting backwards before recall whereas they are not counting before learning. This change of context between learning and recall could account for part of the forgetting in this paradigm. If, as has been found in LTM, context is kept the same for both learning and recall, retrieval should be better than when it differs. To test this hypothesis, *S*s performed the usual Peterson-Peterson task as one condition, and as a second condition, counted backwards *before learning* from the same number for the same length of time used in the retention interval.

---

[1] Support for this research was provided in part by the United States Department of Health, Education, and Welfare, Office of Education, Grant #OEG-3-70-0032 (509) and the National Institute of Mental Health, Fellowship #5 F01 MH41642-02. I wish to acknowledge the assistance of Frank B. Wood, Patricia H. Brown, and Lynn R. Treadway in conducting these experiments.

[2] Requests for reprints should be sent to Philippe Falkenberg, Department of Psychology, Winston Hall, Wake Forest University, Winston-Salem, North Carolina 27109.

Method

Twenty-four Duke undergraduates were paid *S*s. Participation in the experiment was a course requirement. Six retention intervals were used: 3, 6, 9, 12, 15, and 18 sec. One condition (PP69) replicated the Peterson-Peterson paradigm unchanged. The other condition (CON) required *S* to count backwards, from the same number and for the same length of time, immediately before the presentation of the trigram as well as between presentation of the trigram and recall. The *S*s learned consonant trigrams selected from Witmer (1935) with associative values between 4% and 29%. No two successive items contained letters in common, and each trigram occurred equally often over *S*s at each retention interval. A three-digit number from a table of random numbers was paired with each trigram to provide a starting point for backward counting. A green light gave a 2-sec. warning before the trial began and remained lit during the trial. A 1-sec.-long red light signaled when *E* spoke the trigram and later signaled *S* to say the trigram at the end of each retention interval. Intervals were timed with Hunter timers. A black screen hid *E* from *S*. Each *S* was tested under both conditions, and at all retention intervals. In the context condition, CON, another list of trigrams selected from the same source was paired with another set of three-digit numbers, and the same retention intervals were used. (The *S*s counted before learning in this condition, as well as in the retention interval.)

The *S*s went through all trials under one condition before going through the other condition and were separately instructed on the experimental situation before performing under each of the two conditions. The *S*s completed two practice trials before each condition. Each *S* received eight trials for each retention interval for each condition. They had to repeat the trigrams letter by letter after *E* had spoken them, and they had to say three consonant letters for each recall. The experiment was counterbalanced for lists of trigrams, order in which lists were presented, order of conditions, and the order of retention intervals within each condition.

TABLE 1

MEAN PROPORTION CORRECT RESPONSES FOR CON, AND PP69 AT EACH RECALL INTERVAL

Cond.	Retention intervals						
	3	6	9	12	15	18	$\bar{X}$
CON	.90	.85	.82	.71	.72	.64	.78
PP69	.89	.71	.65	.58	.64	.52	.66
Difference between CON and PP69	.01	.14	.17	.13	.08	.12	.12
p (one tail)[a]	.277	.015	.001	.025	.008	.026	.001

Note.—Recall intervals are given in seconds. $N = 24$.

[a] Determined by a signed ranks test.

Results and Discussion

Table 1 shows the mean proportion of correct responses made by the group for every time interval in each condition. Performance in CON is superior to that in PP69 for every time interval, and, except for the 3-sec. interval, these differences are statistically significant. The performance enhancement of CON over PP69 is roughly equal over the time intervals except for the 3-sec. interval. The closeness of the results for the 3-sec. interval for CON and PP69 conditions is a ceiling effect; a major proportion of *S*s scored perfectly at this interval (i.e., eight out of eight trigrams), so little difference could be demonstrated.

Experiment I suggests that if the retrieval context is more similar to the learning context, recall is significantly improved. This would mean that part of the forgetting in the standard Peterson-Peterson paradigm is due to contextual change as well as lengthening retention interval. Alternative explanations for this result, such as easier rehearsal under the context condition, will be considered in Exp. V–VIII.

EXPERIMENTS II AND III

Can a quantitative manipulation of the contextual similarity between learning and recall produce graded results, and can context be varied independently of the retention interval? Experiment II is described first. The *S*s added pairs of digits instead of counting backwards as the context task, and stimuli were presented visually on a memory drum instead of aurally.

Method

The *S*s were 24 Duke University undergraduates, 11 males, 13 females, all paid. Experimental participation was a class requirement. The 144 consonant trigrams *S*s learned had associative values between 4% and 29% (Witmer, 1935). Digits were randomly selected and paired. Every *S* saw the same set of consonant trigrams and digit pairs. Digits and trigram lists were printed by computer and shown on a Lafayette memory drum. Recall latencies were measured with a Standard timer started automatically by the memory-drum paper and stopped

TABLE 2

Experiment II, Means

Measure	Recall errors[a]					Proportion correct	Latencies
	0	33	67	100	$\bar{X}$		
Retention intervals							
5.4	.44	.40	.39	.32	.39	.77	2.25
10.8	.88	.64	.62	.56	.67	.63	2.62
16.2	1.09	.76	.64	.64	.78**	.60**	2.76**
$\bar{X}$	.81	.60	.55	.51**	.61		
Proportion correct	.58	.67	.70	.71**	.66		
Latencies	2.61	2.55	2.48	2.53*	2.54		

Note.—The percent contextual similarity level is designated by the column heads 0, 33, 67, and 100. $N = 24$.
[a] Context × Retention Interval interaction significant, $p < .05$.
* Means differ at $p < .05$, F test.
** Means differ at $p < .01$, F test.

by *S* releasing a push button immediately after recall. A black screen hid *E* from *S*.

The three retention intervals (5.4, 10.8, and 16.2 sec.) and the four context levels (0%, 33%, 67%, and 100%) combined to make 12 conditions. Each trigram was shown in each condition over *S*s. Twelve trials were given for each condition, a total of 144 trials per *S*. These 144 trials were presented to *S*s in the order of a 12 × 12 balanced Latin square so that each condition followed every other condition once. Different *S*s started into the square at different rows so that any practice effect was distributed throughout the square. The *S*s learned and recalled one consonant trigram in each trial. The *S*s added digits saying the totals aloud before learning the trigram as well as before recalling it, to provide similarity of context between learning and recall. To make learning context as similar as possible to recall context, the same additions were used in both cases. Trigrams and digit pairs were presented for .9 sec. In the 100% context level, all the digit addition used in the retention interval was also performed before learning the trigram. If no digits were added before learning, this was the 0% level. A proportion (33% or 67%) of the digits used immediately before recall were added before learning to provide intermediate context levels. The actual numbers of digit pairs involved in each of these context levels was determined by how many were used for each retention interval (e.g., a 5.4-sec. retention interval used six digit pairs, so a 33% context would use the last two pairs before learning; likewise if 12 digit pairs were used in the retention interval, 33% context would use the last four pairs, etc.).

The *S*s were seated before the screen directly facing the memory drum, given the push button to hold, and told how to use it. They were instructed that this was an experiment in STM and forgetting, shown how the trigrams and digits pairs would appear in the memory drum, told to say the trigrams aloud one letter at a time, and to add the digit pairs saying totals aloud. Because of the large number of trials, *S*s were given a 1-min. rest every 12 trials, and 72 trials were run on 2 consecutive days. There was a 9-sec.interval between the presentation of the last digit pair of the retention interval and the beginning of the next trial.

Results and Discussion

In Table 2 all three measures show that as the degree of contextual similarity between learning and recall increases, retrieval performance increases. By far the most sensitive measure of *S*s' performance was the recall error measure. The *S*s could make zero, one, two, or three errors on each trial; an error was defined as saying the wrong letter, e.g., QSX learned, but on recall *S* said QXK, producing two errors. Recall errors are used throughout the remaining experiments (all mean recall errors are pooled over *S*s and trials). In this and all subsequent experiments, all statistically significant main effects and interactions are reported and discussed. The interaction between context levels and recall intervals for errors is significant, $F\,(6, 138) = 2.5$, $p < .05$, and this implies that the total amount of context is having an effect as well as the relative amount between learning and recall. The largest difference between context levels is between the 0% and the 33% context error measures ($p < .01$), but the error difference between the 33% and the 100% context levels is also significant ($p < .025$).

The *S*s took longer to respond when they were wrong than when they were right (wrong $\bar{X} = 3.02$ sec.; right $\bar{X} = 2.27$ sec.).

The latency differences can be explained on this basis: there were more wrong responses, hence longer average latencies, for the long retention intervals and the lowest contextual similarities. Latency also increased with number of errors as did the variability of latency.

The *S*s' learning (or perception) of the trigrams was not perfect. Of the 10,368 consonant letters *S*s were shown, 144 were mislearned as measured by *S*s' responses at the time of trigram presentation, and the error total increased as the number of prior additions increased, indicating that *S*s probably had more difficulty in breaking set for the longer series of additions, and were caught by surprise when the trigram was shown. (When *S*s made learning errors, their recall responses were compared to the wrongly learned response for recall accuracy.) The data show no apparent relationship between learning errors and the context effect. The number of learning errors is far too small to account for the number and relation of retrieval errors. The context effect occurs at retrieval time. The *S*s' number of correct trigrams increased 21% and errors decreased 36% as contextual similarity increased from 0% to 100%.

If contextual similarity between learning and recall is varied quantitatively, *S*s' ability to recall varies proportionately, and context can be manipulated independently of the recall interval. The usual STM forgetting over increased recall intervals was also found.

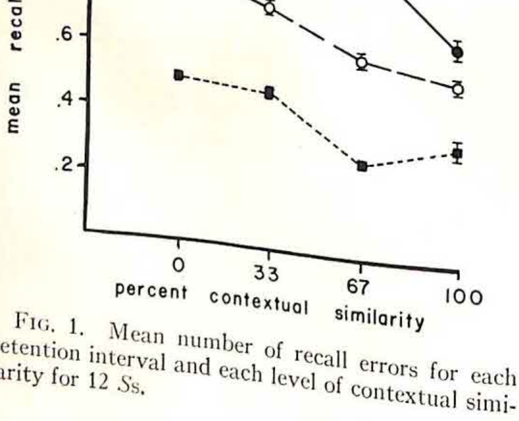


FIG. 1. Mean number of recall errors for each retention interval and each level of contextual similarity for 12 *S*s.

Experiment III was a replication of Exp. II with a new *E*, a better memory drum (Lafayette IBM), a different location, and 12 *S*s from Wake Forest University. Latencies were not recorded. With these exceptions everything was the same as in Exp. II. The results shown in Fig. 1 are similar to those in Exp. II, and the slope of these curves are representative of the context effect in all the studies reported here. The measures of recall performance for different levels of contextual similarity are significantly different, $F\ (3,\ 33) = 6.51$, $p < .005$, as are those for different retention intervals, $F\ (2,\ 22) = 7.38$, $p < .001$, and there is no interaction.

Data from this experiment were further analyzed showing: (*a*) When intrusion errors from the previous trial were examined no systematic relationship between them and the context effect was found. The context effect cannot be explained by proactive interference (PI) from the previous trial. (*b*) The number of errors was least when the recall interval of the previous trial was the same as the present trial. One might speculate that this "sameness" was providing a form of intertrial context effect.

EXPERIMENT IV

The previous three experiments all employed a within-*S* design. Because of the intertrial effects just mentioned, an independent group experiment was done to ensure that the context effect was not some peculiar artifact of the within-*S* design.

Method

The *S*s were 64 undergraduate students, 30 females and 34 males. Participation was a class requirement. Materials and apparatus were the same as in Exp. III, except that only 25 trigrams were used.

The *S*s were randomly assigned to four conditions, 0%, 33%, 67%, and 100% context. A 10.8-sec. retention interval was used. The same list of trigrams was used for each condition, and each *S* received 25 trials. The *S*s were given 2 practice trials and a minute break after every 10 trials.

Results

The mean recall errors for the four conditions were 0%, .82; 33%, .66; 67%, .59; and 100%, .36; and these were significant differences, $F(3, 60) = 3.59$, $p < .025$. These results are extremely close to those obtained in Exp. III for the 10.8-sec. retention interval. Orthogonal tests showed significant differences between the 0% and 33% conditions, and the 67% and 100% conditions. The context effect in the within-*S*s design is not an artifact of the paradigm; because this design is economical of *S*s and more sensitive than an independent groups design we continued to use it in the remaining experiments.

EXPERIMENT V

The previous experiments demonstrated that as you change the amount of contextual similarity between learning and recall for individual *S*s *within a single trial* lasting but a few seconds, as similarity increases so does recall. This has been shown to happen for both visually and aurally presented stimuli, and when the context was backward counting, and when the context was digit addition. What is the mechanism for the context effect?

The simplest explanation would be stimulus generalization, and the form of the data curves would suggest this hypothesis, but alternate explanations exist and must be considered.

The most likely alternate hypothesis is that the addition of context before learning makes rehearsal in the retention interval easier. Since *S*s experienced the same backward counting or digit additions before learning as they did in the retention interval in 100% context, it could be argued that familiarity with the numbers would make them easier in the retention interval, thereby allowing more rehearsal than in the 0% condition. This was tested experimentally.

A second hypothesis involves intertrial interval. Loess and Waugh (1967) have shown that as the intertrial interval increases in STM recall improves. In the first three experiments reported here, this was the case. By adding context before learning, the spacing between the trigram on the present trial and the trigram on the previous trial was increased. Possibly this additional time was allowing "better" learning, giving the context effect, so this was tested.

The third hypothesis tested was that differential activity is producing the context effect. In the 100% context condition, *S*s are busier than in the 0% condition, and possibly this extra activity is "warming them up" so that learning and retrieval are better.

Finally, Keppel and Underwood (1962) propose that the forgetting in this kind of paradigm is caused by proactive interference. The intrusion analysis in Exp. III does not support this explanation for the context effect, although PI clearly occurs; this hypothesis has been tested (Wood, 1971) and PI did not account for the context effect.

Does familiarity with digit addition allow more rehearsal in the 100% context condition? The number of times *S*s saw and added a set of digit pairs within a trial was systematically varied in this experiment at the same time the percentage of contextual similarity was varied.

Method

The *S*s were 36 undergraduate students participating as part of a class requirement. Apparatus, materials, and instructions were the same as Exp. III. Eighty-one trigrams were used. The context levels were 0%, 33%, and 100%. Familiarity with the context was varied by the number of times a particular digit-addition set was repeated in the retention interval. Sets of digit additions were repeated once, twice, or three times in the retention interval; schematically the 0% task looked like this—one repetition, 123456; two, 123123; three, 121212. The context variable was manipulated as before, so that a 33% context level would look like 12 QSX 121212 under a three-repetition situation. Each *S* was tested under all nine conditions of this design, 9 trials each; the 81 trials were presented as a balanced Latin square as before. The *S*s rested 1 min. after each 9 trials, and all trials were done in one session. The retention interval was 10.8 sec., and each trigram and digit pair was presented to *S* for .9 sec.

Results and Discussion

The mean recall errors for contextual similarity were 0%, .66; 33%, .55; and 100%, .52; $F(2, 70) = 5.05$, $p < .01$. Mean recall errors for the number of digit-pair repetitions were 1, .60; 2, .54; and 3, .60; and these differences were not statistically significant. There was no interaction. The context effect was still present, but the number of times *S* experienced the digit additions made no difference in his error scores. The difference between the 0% and 33% context levels was statistically significant ($p < .025$). Since the range of digit-pair repetitions went from zero to five over the context conditions (e.g., in the 100%-three-repetition conditions, *S*s added the same digit pairs three times before learning and three times before recall), *S*s had wide experience with the context task, yet it made no measurable difference to their scores. Differential rehearsal because of familiarity with the context is not a satisfactory explanation for the context effect, but of course differential rehearsal may still be happening for other reasons. This remains to be proven.

EXPERIMENT VI

Method

In Exp. I–III the interval between trigrams varied, being longer for 100% context than for 0%. The average intertrial interval over levels of contextual similarity is controlled in this experiment by varying the spacing from the end of one trial to the beginning of the next; by doing this, the spacing between trigram presentations was held constant for each retention interval. The *S*s were 24 undergraduates participating as part of a class requirement. Apparatus, materials, instructions, and design were the same as Exp. III, except for the control of intertrial intervals.

Results

The upper half of Table 3 shows that the context effect is still present even when the intertrial interval on the average is constant, and the usual increase of errors with increasing retention interval is also present. The magnitude and the form of the context effect data is about the same in this experiment as in Exp. II and III.

TABLE 3

MEAN RECALL ERRORS IN EXPERIMENT VI

Measure	Percent contextual similarity			
	0	33	67	100
Average intertrial interval constant				
Mean recall errors	.93	.76	.67	.67*
Present retention interval	(5.4)	(10.8)	(16.2)	
Mean recall errors	.56	.79	.92*	
Intertrial interval varied[a]				
Mean recall errors	.96	.80	.68	.71*
Previous retention interval	(5.4)	(10.8)	(16.2)	
Mean recall errors	.83	.80	.74	

Note.—$N = 24$. The retention intervals, given in parentheses, are in seconds.

[a] Based on 132 scores per *S*. The first trial in each of the 12 sets was not included.

* Means differ, $p < .01$, *F* test.

Because the retention intervals of the previous trial varied, only the average interval between that trial and the present trial is constant. To check whether this averaging was obscuring any result produced by variable intertrial interval the data were reanalyzed using the previous retention interval as one factor and context as another. This permitted us to look at the data as if the length of the intertrial interval had been deliberately varied; i.e., the intertrial interval between the present and previous trial would vary according to the length of the previous trials' retention interval over the context levels. Under this analysis the context effect remains (lower half of Table 3), and the intertrial interval effect of Loess and Waugh (1967) appears although it is not statistically significant. Loess and Waugh varied their intertrial intervals from 0 to 60 sec., whereas the intervals in this experiment varied from 19.8 to 41.4 sec. Furthermore, the variation of intertrial interval in this last analysis is completely confounded with the variation of the difficulty of the previous trial. In Exp. III, we found that the difficulty of the previous trial did influence performance on the present trial, namely that if the previous retention interval was the same as the present retention interval, performance on the present recall was best. Analysis on this experiment's data confirm this finding. The effect of intertrial interval on recall in this experiment is probably lessened by this confounding and the difference in the range of intertrial intervals used. The significant finding of this study is that the context effect is not appreciably influenced by intertrial interval.

EXPERIMENT VII

Before the influence of activity on the context effect can be examined, a completely different kind of context is needed. So far, context has been an arithmetic task of some kind. Can the context effect be demonstrated using a nonarithmetic contextual task? To answer this question a new context task using words instead of arithmetic was devised.

Method

The *S*s were 18 undergraduates participating as part of a class requirement. With the exception of the context material, apparatus and material were the same as Exp. III. Words chosen from Thorndike and Lorge (occurring at least one per million) were shown to *S*s in the memory-drum slot as context. The *S*s were asked to categorize each word as having something to do with animals or not; in the former case *S*s responded YES, and in the latter case *S*s responded NO, (e.g., *S*s seeing CAT, FORK, ELM, QSX, CAT, FORM, ELM would respond YES, NO, NO QSX, YES, NO, NO). Three context levels—0%, 33%, and 100%—were used, and a digit addition task was included for comparison giving six conditions for the experiment. Each *S* received six trials for each condition organized in a balanced Latin square. The *S*s were instructed in the general nature of the experiment and particularly in both tasks. The retention interval was 10.8 sec. Words, digit pairs, and trigrams were presented for .9 sec. The *S*s rested for 1 min. every 12 trials, and all trials were completed in one session.

Results and Discussion

Both the word-categorization task and the digit-addition task produced the context effect. The mean recall errors for each context condition when words were used as context were 0%, .96; 33%, .74; and 100%, .57; and when digits were used as context, 0%, .78; 33%, .53; and 100%, .54; $F\ (2, 34) = 4.93$, $p < .05$. There was no interaction. The *S*s' recall did not differ on the basis of words or digits; therefore the word-categorization task was equivalent to the digit-addition task in producing response errors. The gradient of context effect scores for the words task was comparable to that found in Exp. II, III, and VI.

EXPERIMENT VIII

Are *S*s being "warmed up" by the additional activity of the 100%, 67%, or 33% context levels in this paradigm? If this is the case, equating the amount of activity over context levels should cause the context effect to vanish.

Method

The *S*s were 36 undergraduate students participating as a part of a class requirement. The materials, apparatus, and design were the same as Exp. V, except for the context. Digit addition was used as context for the three context levels, 0%, 33%, 100%. To equate the amount of activity *S* did in each trial, the spaces before the 0% and 33% were filled with another task. The fill used was: word–categorization, a different set of digits than the ones used in the retention interval, or blanks (no activity) as had been used in previous experiments. The retention interval was 10.8 sec.

The *S*s were instructed in the task as in Exp. V, with additional instructions concerning the context tasks. Both the word-categorization and digit-addition tasks were explained, and *S*s were warned that a mixture of the two might appear on any one trial. The *S*s were given two practice trials on each context task.

Results

Table 4 shows that the context effect remains for each kind of fill. The word and blank fill contribute statistically to the overall significance of the context effect, but the digit fill does not, even though it is in the right direction. The means for each type of fill do not differ significantly, nor do the means for the 100% conditions. The overall number of errors is considerably decreased from previous experiments, but this is understandable since the addition of fill would make the 0% and 33% context levels resemble the 100% level more, thereby reducing error rate. The steepest gradient between levels of contextual simi-

TABLE 4

Experiment VIII, Mean Number of Recall Errors

Type of fill	Contextual similarity			
	0	33	100	$\bar{X}$
Words	.55	.34	.34**	.41
Blanks	.46	.39	.31*	.39
Digits	.40	.37	.35	.38
$\bar{X}$	.47	.37	.34***	.39

Note.—$N = 36$.
* Comparison, $p < .05$.
** Comparison, $p < .01$.
*** Context means differ, $p < .001$, F test.

larity occurs for the word fill. The contrast between the word-categorization task and the digit-addition task might be enhancing the similarity of the digit-addition context making it more salient. Correspondingly, the shallowest gradient occurs for the different digit fill, and this would be expected since the 0% and 33% levels with this fill most closely resemble the 100% level. The gradient of the blank fill context levels most closely resembles the form of the gradients in previous experiments. These results negate the activity or warm-up hypothesis as an explanation of the context effect.

Discussion and Conclusion

Changes in context in an STM paradigm between learning and recall within one trial lasting a few seconds degrade recall performance. In this respect, STM resembles LTM. Three possible explanations for this context effect have been explored experimentally, differential rehearsal, effect of intertrial interval, and the effect of amount of activity in each trial. All have been found wanting.

McGeoch (1942) analyzed causes of forgetting in LTM into four mechanisms: stimulus generalization, change of context, interference, and change of set. This last mechanism is not pertinent to the phenomenon being discussed here. Stimulus generalization may be the mechanism that accounts for the forgetting due to change of context, depending on how the stimulus is defined.

Of the two forms of interference, proactive (PI) and retroactive (RI), the latter may be dismissed summarily. The *S*'s activity in the retention interval was identical for each level of contextual similarity, so it is difficult to see how differential amounts of RI could have produced the context effect. Furthermore, in the 100% situation the interpolated activity in effect was the second trial for that set of digit additions, whereas the 0% set had only one trial. Since two trials should have produced more context learning than one, and RI increases with increased learning of the interfering material, performance should have been worse for the 100% context condition than the 0% condition. The reverse was true.

Although a large amount of PI from the previous trigram was evident, analysis of the intrusions from Exp. III and VI do not show any orderly influence of this interference on the contextual levels. Clearly this is only one source of PI; more remote trials or extraexperimental sources may provide the crucial interference, although Wood (1971) has not found this.

At best the distinction between context and the learned stimulus is arbitrary, and if the context is considered as part of the overall functional stimulus, McGeoch's (1942) first two causes of forgetting can be considered as just one, stimulus generalization. The gradient of recall errors produced as the amount of contextual similarity between learning and recall is quantitatively decreased produces a stimulus generalization decrement much as is found in LTM. If stimulus generalization is the cause of the context effect, as the total amount of contextual similarity increases, the improvement in performance would increase. This occurs in Exp. II, III, and VI, as the recall interval increases. (Note that Peterson and Peterson (1962) found that in recalling lists of paired associates performance decreased if on testing the preceding pair of words in the list was absent. They interpret this to be a stimulus generalization decrement, the stimulus being the list extending through time.)

If the context effect is the result of stimulus generalization, the other half of the curve should be produced if the 100% condition is deviated from by adding more context to it rather than subtracting from it as has been done here. Also, discrimination training should sharpen the gradient, and overtraining in the context material should flatten the gradient. The quantitative context effect is not perfectly clear from these experiments since the effect of adding context is partially confounded by making the retention interval more predictable. Yet, in Exp. VI the retention interval was equally predictable for every trial, but the context effect was still present albeit diminished. More work is needed here.

The context effect is demonstrable *S* by *S*, and not just by averaging large groups. Of the 174 *S*s run in these experiments in within-*S*s designs, 164 showed a difference in error scores between 0% and 100% context, and of these 134 gave fewer errors in the 100% context situation.

Regardless of the mechanism of the context effect, the significance of the above findings is clear. Contextual stimuli do exert a strong influence on people's ability to retrieve from STM if these stimuli are attended to.

REFERENCES

Bilodeau, I. M., & Schlosberg, H. Similarity in stimulating conditions as a variable in retroactive inhibition. *Journal of Experimental Psychology*, 1951, **41**, 199–204.

Carr, H. Maze studies with the white rat: I. Normal animals. *Journal of Animal Behavior*, 1917, **7**, 259–275.

Greenspoon, J., & Ranyard, R. Stimulus conditions and retroactive inhibition. *Journal of Experimental Psychology*, 1957, **53**, 55–59.

Keppel, G., & Underwood, B. J. Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 153–161.

Loess, H., & Waugh, N. C. Short-term memory and inter-trial interval. *Journal of Verbal Learning and Verbal Behavior*, 1967, **1**, 153–161.

McGeoch, J. A. *The psychology of human learning*: An introduction. Longmans-Green: New York, 1942.

Pan, S. The influence of context upon learning and recall. *Journal of Experimental Psychology*, 1926, **9**, 468–491.

Peterson, L. R., & Peterson, M. J. Short term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

Peterson, L. R., & Peterson, M. J. Minimal paired-associate learning. *Journal of Experimental Psychology*, 1962, **63**, 521–527.

Thomas, D. R., & Jones, C. G. Stimulus generalization as a function of the frame of reference. *Journal of Experimental Psychology*, 1962, **64**, 77–80.

Watson, J. B. Kinesthetic and organic sensations; their role in the reaction of the white rat to the maze. *Psychological Monographs*, No. 33, 1907.

Witmer, L. R. The association value of three-place consonant syllables. *Journal of Genetic Psychology*, 1935, **47**, 337–359.

Wood, F. B. Proactive inhibition and context effects in short-term memory. Unpublished master's thesis, Wake Forest University, 1971.

(Received October 27, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 48–54

CONTINUOUS JUDGMENTS OF WORD FREQUENCY AND FAMILIARITY[1]

IAN BEGG[2] AND EDWARD J. ROWE[3]

University of Western Ontario

Three experiments are reported in which *S*s made frequency or familiarity judgments of meaningful words presented in a continuous sequence. In Exp. I, the judged familiarity of 100 words was found to be logarithmically related to exposure frequency, with high-imagery items receiving higher ratings than low-imagery items. Experiments II and III showed that judgments of situational frequency accurately reflected true frequency and were little affected by the imagery value or preexperimental frequency of the items. A model relating frequency judgment processes to recognition memory was presented to account for the experimental findings.

This study examined judged frequency and familiarity of words as a function of their frequency of presentation. In Exp. I, the relationship between exposure frequency and judged familiarity was explored, using a variant of the continuous recognition memory paradigm (cf. Shepard & Teghtsoonian, 1961). The *S*s were presented with a list of words in which individual items appeared varying numbers of times. At each presentation, *S*s rated their familiarity with the word in question. In Exp. II and III, the words were presented as in Exp. I, but *S*s judged the number of prior occurrences of each word.

Besides the obvious purpose of relating judged familiarity and frequency to the number of occurrences of an item, an attempt was made to assess the relationship of the image-arousing capacity (*I*) and Thorndike-Lorge frequency (F) of the words to the above mentioned measures.

EXPERIMENT I

Previous studies have shown that rated familiarity is a negatively accelerated function of exposure frequency, although there has been some question about the precise form of the function. Noble (1954) found that the familiarity of dissyllables was a hyperbolic function of the frequency of presentation of the items, while Arnoult (1956) reported that the relationship for nonsense shapes was logarithmic. The primary difference between the two functions is that the hyperbolic curve approaches a horizontal asymptote, whereas the logarithmic function does not. Smith and Dixon (1971) varied the Thorndike-Lorge F value of meaningful words in a familiarity-judgment task and found that the ratings were more closely approximated by a logarithmic than a hyperbolic function. They concluded that "familiarity is a comparative quality of words which depends primarily upon an individual's ability to accurately estimate his relative frequency of encounter with various stimuli . . . it appears to be a process which continues far beyond the frequency limits suggested by Noble's work [pp. 280–281]."

The present experiment examined the familiarity-frequency relation for meaningful words, with some procedural modifications. First, *S*s made familiarity judgments after each word in the list was exposed, whereas Noble's (1954) and Arnoult's (1956) *S*s made their ratings after the entire list had been presented. The present design should permit an examination of the trial-by-trial change in familiarity of real words as a function of the number of presentations. Second, all three of the studies cited above (Arnoult, 1956; Noble, 1954; Smith & Dixon, 1971) used rating scales for *S*s' judgments. It is possible that the use of such scales imposes an upper limit on the responses that is not a true reflection of changes in subjective familiarity. Therefore, *S*s in the present experiment used a magnitude estimation procedure, in which they used any numbers they wished as ratings. Additionally, words varying in *I* were used as stimuli in order to determine whether the relationship between frequency and familiarity would be similar in the two classes or words.

---

[1] The research was conducted while both authors held National Research Council of Canada Postgraduate Scholarships at the University of Western Ontario. Preparation of the manuscript was financed in part by National Research Council of Canada Grant A8122 to the first author. Thanks are due to Allan Paivio for his encouragement and to Christine Bills, who helped in collecting the data.

[2] Requests for reprints should be sent to Ian Begg, Department of Psychology, McMaster University, Hamilton, Ontario, Canada.

[3] Now at the Department of Psychology, Memorial University, St. John's, Newfoundland, Canada.

Method

Subjects.—Six introductory psychology students at the University of Western Ontario served as *S*s as part of a course requirement.

Materials and procedure.—Fifty High-*I* and 50 Low-*I* nouns were selected from the Paivio, Yuille, and Madigan (1968) norms, with respective mean *I* ratings of 6.39 and 3.04 on a 7-point scale. All words had F values of A or AA. In constructing the list, 10 randomly selected High-*I* and Low-*I* nouns were each printed on 17 computer cards; 10 High-*I* and Low-*I* nouns each on 9 cards; and 10 each on 5, 3, and 1 cards. The 700 cards thus produced were then arranged in 28 sets of 25 cards each. Each set was thoroughly shuffled, and the 28 sets were put together such that cards containing the same word appeared in successive sets. The pile in which the initial presentation of an item occurred was randomly determined, with the restriction that enough adjacent piles remained to contain all the repetitions of a particular word. An average of 25 intervening items appeared between successive occurrences of a given word within the list.

The complete deck of 700 cards was placed in a pile in front of *S*, who went through the cards 1 at a time at his own rate. The *S*s were required to read each word aloud and assign to it a number to represent how familiar they were with the word in question. They were instructed to use any numbers they wished, but to maintain a relationship between the numbers they used and their subjective feelings of familiarity associated with the words. The procedure took about 30 min.

Results and Discussion

Since *S*s used purely subjective scales to estimate familiarity, each *S*'s responses were converted to *z* scores, and mean *z*s were calculated across *S*s. The means increased linearly with $\log_{10}$ frequency of presentation (Fr), with judged familiarity (F′) = .521 log Fr − .225, $r = .968$ for the High-*I* nouns, and F′ = .478 log Fr − .423, $r = .962$ for the Low-*I* nouns. As can be seen from Fig. 1, the functions were roughly parallel, with High-*I* words being rated as as more familiar than Low-*I* ($p < .001$ by a sign test).

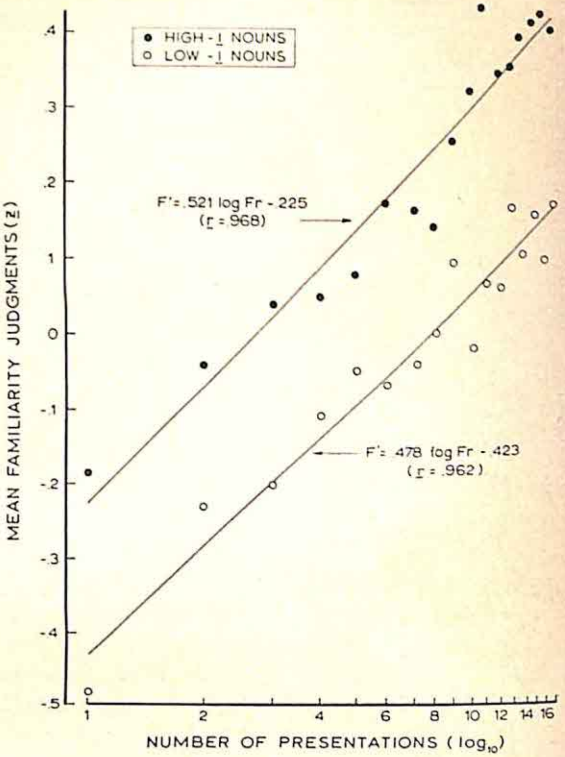


FIG. 1. Mean rated familiarity of High- and Low-*I* nouns as a function of the frequency of presentation.

The familiarity-frequency relationship for meaningful words thus appears to be described quite well by a logarithmic function, as found by Arnoult (1956) for nonsense shapes. The results also compare very favorably with those of Smith and Dixon (1971), who related familiarity to F. While the logarithmic fit to their data was not as good as the fit obtained here, the slopes of the functions are very similar, being close to .50 in both cases. Thus both preexperimental and experimental frequency seem to be related to familiarity in the same way, and furthermore, this relationship is insensitive to a number of differences in procedure.

Finally, the logarithmic function describes the data for High- and Low-*I* nouns about equally well, although for some reason the High-*I* nouns were judged as being more familiar than the Low-*I* nouns.

EXPERIMENT II

Experiments II and III were designed to examine *S*s' ability to estimate the experimental frequency of occurrence of meaningful words. Hintzman (1969) found *S*s' frequency judgments increased linearly with log Fr and suggested, after Peterson (1966), that the strength of the long-term memory (LTM) trace corresponding to an item was a function of the frequency of presentation of the item. Theoretically, frequency judgments could thus provide an index of the strength of the LTM trace, an assumption which has been implicit in the work of other investigators (Underwood, 1969; Williams & Underwood, 1970). The *S*s in Hintzman's experiment made their judgments after the entire list of words had been presented. They were not instructed prior to presentation that frequency judgments would be required, although the latter variable does not seem to be important, since Underwood and Freund (1970) essentially replicated Hintzman's findings using an intentional instruction procedure.

Before accepting Hintzman's (1969) suggestion that LTM trace strength is a *logarithmic* function of Fr, it is important to isolate the locus of *S*s' failure in accuracy at higher and lower levels of Fr. It could be that *S*s cannot keep track of the number of times each item has occurred during trial-by-trial presentation, in which case the logarithmic relationship would reflect short-term memory (STM) influences. Alternatively, given that *S*s can keep an accurate count of the number of times each item has occurred during presentation, the mechanism underlying frequency estimation might change systematically in the interval between the last presentation of an item and the time at which judgment is requested, thus implicating the influence of LTM processes alone. As a test of these alternatives, the continuous-judgment paradigm was used in the present experiment to study the trial-by-trial incrementation of frequency in detail, with *S*s judging the frequency of each item immediately following its presentation. In the present usage, STM is defined as the interval between presentations of an item, while LTM is defined as the cumulative interval over repeated presentations.

Method

Subjects.—Six *S*s were selected from the same population as in Exp. I.

Materials and procedure.—The materials from the first experiment were used. Each *S* went through the deck of 700 computer cards at his own rate. The words were read aloud, and *S* gave a number to represent how many times each item had occurred previously in the list. That is, the correct response on the initial presentation of an item was "zero," on the second presentation "one," and so on.

Results and Discussion

Median judgments were calculated across the six *S*s for each level of Fr. These are plotted separately for High-*I* and Low-*I* nouns in Fig. 2, with the line in the figure representing perfect accuracy. The judgments showed no tendency toward an asymptote and were, in fact, almost perfectly linearly related to Fr, $r = .997$ for both types of items. The difference between High-*I* and Low-*I* nouns was not as large as in the familiarity ratings of Exp. I, but High-*I* nouns were rated as having occurred more frequently at 12 of the 14 frequency levels where there were no ties ($p < .05$ by a sign test).

The function relating frequency judgments to frequency of presentation thus appears to be linear in a short-term task, rather than logarithmic, as found by Hintzman (1969) in a longer-term task. This suggests that the logarithmic function is a result of the delay imposed between item presentation and judgment, i.e., LTM influences.

The present findings also allow an evaluation of Smith and Dixon's (1971) suggestion that familiarity ratings reflect the ability of *S*s to estimate exposure frequency accurately. If this were true, the form of the frequency function in Exp. II should be similar to the familiarity function of Exp.

I. That is, the relationship between judged and actual frequency should be logarithmic and High-*I* items should be judged as having occurred more frequently than Low-*I* items. While High-*I* words did receive higher frequency estimates in the present experiment, this difference was much less pronounced than for the familiarity judgments; in addition, the function relating judged frequency to Fr is clearly linear. These findings indicate that familiarity judgments are not determined solely by the ability to estimate Fr accurately.

It is possible to examine recognition memory performance by considering only the responses to words presented for the first or second time, with any nonzero response to an item on its first presentation being a false alarm, and any nonzero response to a second presentation being a hit. High-*I* nouns were associated with higher hit rates, .937 versus .850, lower false alarm rates, .113 versus .163, and, consequently, higher corrected hit rates than Low-*I* nouns, regardless of whether high-, or, low-threshold correction assumptions are involved. The hit data are based on 240 observations and the false alarms on 300. The data are presented with no assessment of the reliability of the differences. Experiment III was partially a replication of Exp. II, enabling the reliability of the results to be assessed directly.

Since recognition memory favored High-*I* nouns, the probability of an exactly correct numerical judgment should also favor these items, although such probabilities should approach zero for both levels of *I* as Fr increases. For both word types, accuracy (Pc) declined linearly with log Fr, with $\text{Pc} = -.630 \log \text{Fr} + .776$, $r = -.934$ for the High-*I* nouns and $\text{Pc} = -.618 \log \text{Fr} + .721$, $r = -.959$ for Low-*I* nouns. The two functions do not differ much in slope, with the High-*I* nouns having about a 5% advantage throughout most of the range of frequencies. Overall, Pc was greater for High-*I* than Low-*I* nouns in 13 of 17 cases ($p < .05$).

Also, since recognition favored High-*I* over Low-*I* words, it would be expected that the proportion of judgments which underestimate true frequency (Pu) would be higher in the case of Low-*I* items. For both levels of imagery, Pu increased linearly with log Fr, $\text{Pu} = .275 \log \text{Fr} + .041$, $r = .946$ for High-*I* nouns and $\text{Pu} = .333 \log \text{Fr} + .090$, $r = .911$ for Low-*I* nouns. The correlations were calculated without considering the first presentation of an item, since it was not possible for *S*s to underestimate that occurrence. Therefore, the *y*-intercept values of the regression lines are meaningless. A more meaningful measure is the *y* ordinate of the regression line at $\text{Fr} = 2$ ($\log \text{Fr} = .30103$), which yields .124 for the High-*I* and .190 for the Low-*I* nouns. To the extent that underestimation reflects forgetting, Low-*I* nouns were forgotten more often throughout the task, i.e., in 15 out of the 16 comparisons ($p < .05$).

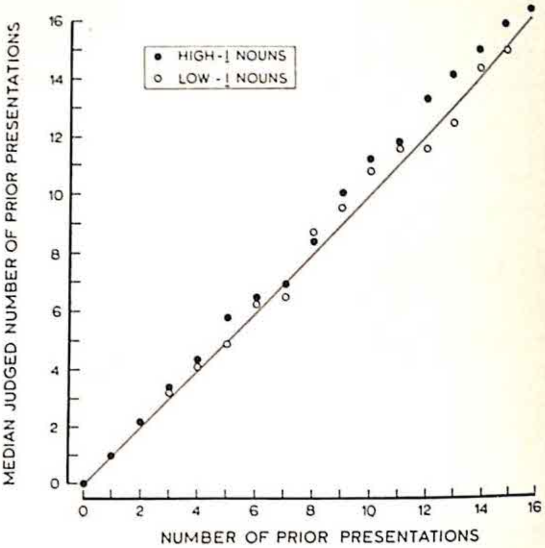


FIG. 2. Actual and judged number of prior presentations of High- and Low-*I* nouns.

EXPERIMENT III

In Exp. III, *I* and F were varied factorially in the continuous judgment task in an attempt to test both the replicability and the generalizability of the results obtained in Exp. II. Considering High-F words alone, several predictions are clear. First of all, High-*I* and Low-*I* nouns should not differ appreciably in median frequency judgments, which should approximate true frequencies. Second, *I* should be positively related to hit rates and negatively related to false alarms. Third, the proportion of exactly correct judgments should decline linearly with log Fr, with the regression lines for High-*I* nouns having higher intercepts than the lines for Low-*I* nouns. Finally, the proportion of frequency judgments which underestimate true frequency should increase linearly with log Fr for both levels of *I*, with the Low-*I* nouns underestimated more often that High-*I* nouns. These effects attributed to *I* should also generalize to Low-*F* words.

If the effects obtained for imagery in Exp. II are a reflection of the recognizability of the words, analogous results would be expected for the F variable. That is, since F is negatively related to recognition memory (e.g., Allen & Garton, 1968; Shepard, 1967), the positive effects of *I* obtained in Exp. II should be mirrored by negative effects of F. Thus, hit rate should be a negative function and false alarms a positive function of F; the proportion of exactly correct judgments should be describable by a regression line having a higher intercept for Low- than for High-F words; and underestimates should be higher for High-F words than for Low-F words.

In sum, Exp. III was an attempt to (*a*) replicate the effects of *I* obtained with High-F nouns in Exp. II, (*b*) extend the generality of the effects to Low-F nouns, and (*c*) conceptually extend the results to F, a variable having opposite effects on recognition memory, thus allowing the conclusion that recognition memorability is a necessary precursor of frequency incrementation.

Method

Subjects.—Six *S*s were chosen from the same pool as those in the first two experiments.

Materials and procedure.—Nouns varying factorially over two levels of *I* and F were selected from the Paivio et al. (1968) norms, with 25 nouns representing each combination of high and low I and F. The High-*I* nouns had mean *I* ratings of 6.35 and 6.32 for the High-F and Low-F words, respectively, while the corresponding ratings for the Low-*I* nouns were 3.06 and 3.05. All High-F nouns had F values of A or AA, and the Low-F nouns had values of less than 10 per million by the Thorndike-Lorge G count. Five randomly selected words of each type were printed on 16 computer cards five each on 8 cards, and five on each of 4, 2, and 1 cards. The 620 cards were arranged in 24 sets of 25 cards each and 1 set of 20 cards, with cards containing the same noun placed in consecutive sets as before. The sets were thoroughly shuffled and put together, with the set of 20 cards last. The instructions and procedure were identical to those of Exp. II, with each *S* serving for 620 trials.

Results and Discussion

The median frequency judgments, calculated across *S*s, are plotted in Fig. 3, with the line in the figure representing perfect accuracy. The correlations between actual and judged frequencies for the four word types were all greater than .98. High-*I*–High-F nouns were rated as occurring more frequently than Low-*I*–High-F nouns in 11 out of 14 non-tied comparison ($p = .06$ by a sign test) and, as in Exp. II, the absolute differences were not large. These same Low-*I* nouns were rated as having occurred less frequently than their Low-F counterparts in 10 out of 14 comparisons. Again, the differences were small. It is thus safe to conclude that running frequency estimates, on the average, approximate true frequencies very closely and are not affected to any great extent by item characteristics.

The recognition memory results, calculated as in the previous experiment, closely replicated those findings. For High-F nouns, *I* was positively related to hits, .867 versus .950, and negatively related to false alarms, .180 versus .073, exactly as in Exp. II. The direction of each of these effects extended to Low-F nouns, with hit rates of .967 and .983, and false-alarm rates of .080 and .033, although the difference in each case was considerably smaller. The effects of F, as expected, were the reverse of the imagery effects. Within High-*I* nouns, F was negatively related to hits, .983 versus .950, and positively related to false alarms, .033 versus .073. The same directional results obtained with Low-*I* nouns, with hits of .967 and .867, and false alarms of .080 and .180. The positive effect of *I* and the negative effect of F were about equal in magnitude. The effect of *I* was most pronounced in

High-F words, while F was most effective in Low-*I* words.

If the results of the analysis of proportion correct in the previous experiment reflect memorability of the nouns, the positive effects of *I* on the intercept of the regression line should occur in the present experiment at both levels of F, and there should also be a corresponding negative effect of F. The results were in accord with prediction. All correlations between Pc and log Fr were in excess of −.95. The highest intercept value was found for High-*I*–Low-F (Pc = −.858 log Fr + .958), followed in order by High-*I*–High F (Pc = − .776 log Fr + .887), Low-*I*–Low-F (Pc = − .721 log Fr + .841), and Low-*I*–High-F (Pc = − .630 × log Fr + .709).

With underestimates, more memorable items should be associated with lower intercept values (at log Fr = .30103) than less memorable items. All correlations between Pu and log Fr exceeded .93, and the order of corrected intercepts was the reverse of those in the Pc data, with Low-*I*–High-F being highest, then Low-*I*–Low-F, High-*I*–High-F, and High-*I*–Low-F. The respective intercepts were .167, .056, .043, and .039, with regression equations of Pu = .486 log Fr +. 021, Pu = .532 log Fr − .104, Pu = .565 log Fr − .127, and Pu = .581 × log Fr − .136. As in the Pc data, the slopes were generally converging.

General Discussion

In this section, the results will first be iterated in brief, followed by a consideration of more general issues. The first experiment contributed one major result: namely, that judged familiarity of words is a logarithmic function of their exposure frequency. That is, the frequency-familiarity relationship remains logarithmic regardless of whether (*a*) frequency is defined by frequency of experimental exposure or frequency of occurrence in print (cf. Smith & Dixon, 1971), (*b*) familiarity is defined by an open-ended magnitude estimation procedure or a fixed-end rating-scale judgment (cf. Arnoult, 1956; Smith & Dixon, 1971), or (*c*) the stimuli are real words or nonsense shapes.

The second and third experiments provided a number of results. In the first place, the median judgments of frequency were almost identical to the true frequencies for all word classes. Hence, all repetitions of the items, on the average, produced increments in whatever internal state corresponds to exposure frequency. It is interesting to note that the familiarity ratings from Exp. I do not show a similar linear increase, suggesting that running familiarity ratings are not determined directly by subjective frequency.

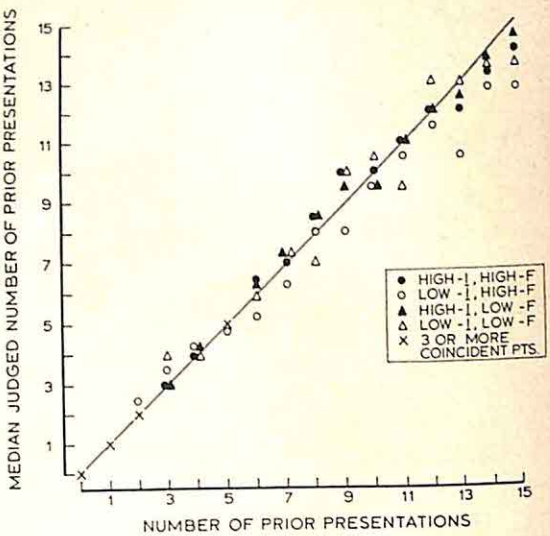


Fig. 3. Actual and judged number of prior presentations of High- and Low-*I* and F nouns.

A class of results obtained in the second and third experiments can be summarized in terms of their relationships to the recognizability of the words. It is already known that *I* and F have opposite effects in recognition memory experiments. Consequently, variables which are coincident with recognition should have opposite relations with *I* than with F. A number of such variables were discussed in the Results section of Exp. III.

The results raise questions concerning the relationship between frequency judgments and the memory trace or, more generally, how repetitions are coded in memory. Assume that the initial presentation of an item establishes a memory trace. Hintzman and Block (1971) discussed two possibilities: whether subsequent presentations of that item cause the trace to become stronger or whether they cause new traces to be formed. They decided on the basis of their data that the second possibility is more likely. Their argument will not be duplicated here but will be for the most part accepted. The question then boils down to one of how the new trace is related to the initial one. The process of continuous frequency judgments might best be conceived as comprising three stages. In the first stage, *S* must recognize a given presentation of an item as being a repetition. Following this, a paired associate is formed in which the stimulus term is the item and the response term is a number (probably "one" on the second presentation). On each subsequent presentation of the item, *S* must be able to retrieve the pair consisting of the item and its number, increment the numerical associate by one, and read out the new response term. Anecdotally, some *S*s in Exp. II and III, upon a presentation of an item, would remark "What did I say last time?" Of course, the proposed model needs explicit testing before it can confidently be applied as an explanation of the frequency judgment process.

Finally, the results of the experiments reported here have implications for verbal discrimination learning, where situational word frequency has been viewed as the dominant attribute affecting performance (Ekstrand, Wallace, & Underwood, 1966; Underwood & Freund, 1970). It has been argued that the comparative ease of learning High-*I* over Low-*I* word pairs in this task is not explainable in terms of frequency theory (e.g., Rowe & Paivio, 1971). The fact that the judged frequency of these two word types in Exp. II and III of the present study was not appreciably different concurs with this point of view. Even though High-*I* words were given significantly higher median judgments than Low-*I* words in Exp. II, the absolute differences were small compared to differences in performance typically observed in discrimination learning and were actually not significant in Exp. III. Also, on the first two or three trials in the present study, there was no difference attributable to imagery in the median frequency judgments. In the verbal discrimination paradigm, the imagery effect is typically present on the first or second trials. Thus, to the extent that frequency judgments reflect the operation of frequency processes in verbal discrimination learning, we conclude that the effect of imagery in that task is not amenable to an interpretation in terms of situational frequency.

REFERENCES

ALLEN, L. R., & GARTON, R. F. The influence of word-knowledge on the word-frequency effect in recognition memory. *Psychonomic Science*, 1968, **10**, 401–402.

ARNOULT, M. D. Familiarity and recognition of nonsense shapes. *Journal of Experimental Psychology*, 1956, **51**, 269–276.

EKSTRAND, B. R., WALLACE, W. P., & UNDERWOOD, B. J. A frequency theory of verbal-discrimination learning. *Psychological Review*, 1966, **73**, 566–578.

HINTZMAN, D. L. Apparent frequency as a function of frequency and the spacing of repetitions. *Journal of Experimental Psychology*, 1969, **80**, 139–145.

HINTZMAN, D. L., & BLOCK, R. A. Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, 1971, **88**, 297–306.

NOBLE, C. E. The familiarity-frequency relationship. *Journal of Experimental Psychology*, 1954, **47**, 13–16.

PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology* 1968, **76**(1, Pt. 2).

PETERSON, L. R. Short-term verbal memory and learning. *Psychological Review*, 1966, **73**, 193–207.

ROWE, E. J., & PAIVIO, A. Word frequency and imagery effects in verbal discrimination learning. *Journal of Experimental Psychology*, 1971, **88**, 319–326.

SHEPARD, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 156–163,

SHEPARD, R. N., & TEGHTSOONIAN, M. Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, 1961, **62**, 302–309.

SMITH, R. C., & DIXON, T. R. Frequency and the judged familiarity of meaningful words. *Journal of Experimental Psychology*, 1971, **88**, 279–281.

UNDERWOOD, B. J. Some correlates of item repetition in free recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 83–94.

UNDERWOOD, B. J., & FREUND, J. S. Relative frequency judgments and verbal discrimination learning. *Journal of Experimental Psychology*, 1970, **83**, 279–285.

WILLIAMS, R. F., & UNDERWOOD, B. J. Encoding variability: Tests of the Martin hypothesis. *Journal of Experimental Psychology*, 1970, **86**, 317–324.

(Received October 29, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 55–62

RECOGNITION MEMORY FOR PERCEPTUALLY SIMILAR PICTURES IN PRESCHOOL CHILDREN [1]

ANN L. BROWN [2] AND JOSEPH C. CAMPIONE

Children's Research Center, University of Illinois

A continuous recognition memory task was employed in three experiments with preschool children. On test trials, the stimuli were from one of three categories, new, old-identical, or old-similar. The latter category consisted of pictures of a character seen previously (hence old) but now shown in a similar, but different, pose. Experiment I investigated short-term (within a session) retention, and Exp. II and III investigated retention over longer periods (up to 7 days). In general, the data from Exp. I and III indicated that *S*s were able to discriminate identical and similar recurrences of a central character even up to a 7-day retention interval, demonstrating that considerable detail of the original picture was both stored initially and retained for an extended period. In Exp. II, *S*s consistently responded correctly on old-similar items but made more errors on old-identical items, with this difference increasing with longer retention intervals. This divergence was attributed to a response bias; i.e., with increasing retention interval, *S*s' tendency to respond SIMILAR rather than IDENTICAL increased. The results obtained in Exps. III supported this interpretation.

Recent studies of recognition memory suggest that the ability to identify repeating pictures is well developed in preschool children. Children's recognition of items repeating within a session (Brown & Scott, 1971; Corsini, Jacobus, & Leonard, 1969) and over retention intervals up to a month (Brown & Scott, 1971) is comparable to that reported for adults (Nickerson, 1965, 1968; Shepard, 1967). In these experiments, the pictures chosen have generally been as distinct as possible. For example, in the child studies, the stimuli have all been pictures which featured a central character, with background details minimized. Further, these studies tended to use only one exemplar of each item within a category (e.g., only one lion in the animal category). As such, these studies may not provide a realistic estimate of the children's recognition capacity. For example, Reese (1970) has suggested that memory in young children is in terms of an iconic representation, the details of which fade with time, ("leveling"), leaving only a figurative image (dog, lady, lion) stripped of idiosyncratic features. If such were the case, the use of pictures emphasizing distinct central characters would not provide a realistic estimate of children's capacity. Also, it may be the case that the use of simple pictures stressing one focal character or object aids the young child in abstracting the critical features of the item for later comparison. To investigate either possibility, it would be necessary to investigate children's performance in a situation where the pictures entered in the stimulus pool were not as distinct from each other.

Little is known about the effect of picture similarity and recognition accuracy even in adults. However, studies of recognition of isolated words and connected discourse have attempted to specify stimulus similarity in terms of the critical features of storage. Possible analogues for picture recognition may be found in this literature. One method of determining similarity among input words has been to discover the type of words which lead to false recognition that a new (distracter) item is old. Common associates of target words have produced significantly more false recognition responses than control words in both children (Hall & Ware, 1968) and adults (Davis, 1967; Underwood, 1965). Anisfeld & Knapp (1968) explain their finding of increased false recognition of synonyms of target words by suggesting that the word as a whole is not the unit of storage. The word is stored as a complex of its constituent features, such as semantic, syntactic, phonological, or orthographic features. These features are stored independently and are subject to differential decay rates. False recognition occurs when a match is made between a distracter word and a target word which are alike on one or more, but not all, features. An analysis of which feature similarities lead to a mismatch provides information concerning which critical features of the word have been stored and which fade with time. There is some evidence that for long-term storage, verbal material is encoded according to its semantic features. Studying recognition of sentences, Sachs (1967) found that changes in the semantic features of a sentence continued to be recognized over time, while syntactic changes were not as well remembered.

---

[1] This research was supported by United States Public Health Service Grants HD 05951 and M 1099. The authors would like to thank the staff and children of the Hans Christian Andersen Montessori School, and the Children's World Day Care Center, both of Tolland, Connecticut.

[2] Requests for reprints should be sent to Ann L. Brown, Children's Research Center, University of Illinois, Champaign, Illinois 61820.

If words are stored according to certain critical features, are pictures similarly stored in terms of features or as wholes? Again, this question can best be asked for pictures, as it is for words, by varying the similarity of the distracter items to the target item and analyzing any subsequent false recognition responses. In this paper, one form of similarity was investigated. The question was: Are pictures stored in terms of some gross feature such as the central character, or are details concerning the action of that central character also available? The items used were, therefore, pictures in which the central character was always a person or an animal with background details minimized. Test pictures consisted of either an identical test where the same character appeared in the identical pose or a similar test where the same character (same clothes, same colors, etc.) appeared in a similar pose "doing something different." The action of the character had changed. An immediate within-session test of the ability to distinguish between the two types of items was provided in Exp. I. In Exp. II and III long-term retention of these critical features of the pictures was tested.

Experiment I

Method

Subjects.—The *S*s were 15 preschool children, 9 boys and 6 girls. They ranged in age from 37 to 62 mo., with a mean of 52. Two further 3-yr.-olds were unable to understand instructions and were dropped from the experiment.

Stimulus materials.—Colored pictures of either people or animals were cut from children's books and pasted in the center of 35.6 × 15.2 cm. black posterboard cards. There were two experimental stacks of 110 cards each, consisting of 50 duplicates and 10 single (filler) items. Half of the duplicates were identical pairs, and the remaining duplicates were similar pairs (i.e., the same character occurred in both pictures of a pair but in different poses). A further stack of 20 cards, consisting of five identical pairs and five similar pairs, was used in pretraining.

Procedure.—The two experimental stacks were ordered so that the lag between the presentation of the first and second item of a pair was either 0, 5, 10, 25, or 50 intervening pictures. In each stack, there were 10 items recurring after each lag; of these, 5 items were identical to the first member of its pair and the remaining 5 items were similar. Each *S* was tested on both stacks and thus provided 10 observations for identical and 10 for similar items at each of the five lags. The pictures were further ordered so that within each successive block of 20 cards an equal number of Lag 0 and 5 items occurred. Tests of the remaining lags were distributed as evenly as possible throughout the series, with the necessary restriction that a greater proportion of the long lag tests occurred in the latter part of the series. An attempt was made to avoid runs of more than three new items in a row, and the proportion of old to new items was kept approximately stable across blocks. The pretraining stack (20 cards) was ordered so that the second member of both identical and similar pairs occurred after a lag of either 0, 2, or 5 intervening items.

Each *S* was run individually. Day 1 consisted of the pretraining stack plus one of the experimental stacks. Both of these stacks were placed face down in front of *S*. The *E* turned the cards over one at a time and waited for *S*'s response. Approximately 5 sec. were required to view and respond to each picture. On each trial, *S* was instructed to say YES if he had seen the character before and NO if he had not seen the character before. If he responded YES, he was further required to say whether he had seen the character doing "exactly the same thing" or "something different." After the first few trials, the majority of *S*s spontaneously provided the additional information (e.g., "Yes, I've seen him before, but now he is running," or "Yes, I saw him just like that."). As each repeat occurred in the pretraining stack, *E* returned to the first occurrence of that item and said, "Yes, It's exactly the same," or "No, look, he(she) is doing something different."

Following pretraining, *S*s were assigned randomly to one of the two experimental stacks. Here *E* no longer indicated that *S*'s response was correct or incorrect by returning to the first occurrence of a picture. Knowledge of results was not given, but *E* periodically indicated that *S* was "doing fine." On the following day, the second stack of cards was presented, following a review of the pretraining cards. On completion of each series, *S* chose a small reward item from a collection of 10¢ toys.

Results

On repeating items, a correct response required that *S* indicate (by saying YES) that he had seen the character before *and* correctly class it as identical or similar to the previous occurrence of that character. The proportions of correct responses for the identical items were .99, .99, .95, .91, and .89 for lags of 0, 5, 10, 25, and 50, respectively. The corresponding figures for the similar items were .99, .98, .93, .95, and .91. A 3 (Lag) × 2 (Test Type) × 15 (*S*s) analysis of variance was conducted on the number of errors on a lag of 10, 25, and 50 items. Lags 0 and 5 were excluded, since there was virtually no variance in these cells. None of the effects were reliable, although the lag effect approached significance, $F(2, 28) = 2.80$, $.10 > p > .05$, indicating a tendency toward increasing error rates with longer lags. The effect of lag was further investigated by comparing performance on Lag 0 and 5 items with performance on Lag 25 and 50 items. All of the 15 *S*s were more accurate on the shorter lag items ($p < .01$, sign test). The false-alarm rate, the proportion of new items classed as old, was extremely low, with only 13/1800 such errors occurring.

Discussion

The data are clear in demonstrating that considerable detail of the pictures was stored on the original presentation of an item. Even with lags of 50 items, the probability that *S*s could identify the second member of the pair as either identical or similar to the initial member remained around .90. Thus, even with 50 intervening pictures, *S*s were not only able to recognize that they had seen a certain character before, but were capable of reporting whether or not he (she) was doing the same thing. However, it is possible that the high levels of performance were due to the fact that the pretraining and use of a continuous recognition procedure "set" *S*s to look for details of the pictures. That is, the pretraining warned *S*s that small changes in detail would be important. Specifically, they were warned that action changes would be important, as they were informed that on some occasions the character would be seen "doing something different." Experiments II and III avoid this problem by presenting all the pictures initially before *S*s are given information regarding the importance of such changes.

Given that there was some evidence of a decline in performance at the longer lags, Exp. II was conducted to determine whether this resistance to interference from similar items would be maintained with longer retention intervals, up to 7 days. If Reese's (1970) suggestion is correct, we would expect that the ability to distinguish similar from identical items would disappear with longer retention intervals.

EXPERIMENT II

Method

Subjects.—The *S*s were 21 preschool children. Three children were dropped from the experiment, 1 for failing pretraining and 2 for failure to reach criterion on an immediate retention test. The remaining 18 *S*s, 9 boys and 9 girls, ranged in age from 39 to 59 mo., with a mean of 51.

Stimulus materials.—The stimulus materials were similar to those used in Exp. I. Colorful pictures of people and animals were cut from children's books and pasted on black posterboard cards. There were 80 pairs of pictures, 40 identical and 40 similar, and 80 individual pictures, for a total of 240 pictures.

Procedure.—All testing was done individually, with the experiment being conducted in two phases. The *S*s were initially shown an inspection stack of 80 pictures, one from each of the pairs of pictures available. They were told to look at them carefully, as they would be required to remember them later. Immediately following the inspection series, *S*s were presented a 20-item pretraining stack consisting of 10 old items (5 identical and 5 similar to a picture that had occurred in the inspection series) and 10 new characters which had not occurred before. As in Exp. I, *S*s were trained to respond YES to both identical and similar items and NO to new items, and to indicate whether an item was identical or similar following a YES response.

Following pretraining, an immediate test of 20 items was given to ensure that the pretraining task was sufficient to introduce *S*s to the rules of the game. This stack also consisted of 10 old items (5 identical and 5 similar to an item in the inspection stack) and 10 new items. The *S*s were required to respond correctly 80% of the time to each of the three types of items before being permitted to continue. The *S*s were given 10¢ toys at the end of the immediate test and told that they would return later to see how well they remembered the pictures.

The *S*s were randomly assigned to one of three retention intervals: 2 hr., 1 day, or 7 days. At the designated interval, each *S* was shown the pretraining stack again and reminded of the rules of the game. Following this, the retention stack was presented. This consisted of 120 pictures, 60 old items (30 identical and 30 similar) and 60 new items. Again, *S*s were required to respond YES to all old items and NO to new items. If they responded positively, they were further required to indicate whether the item was identical or similar (exactly as in Exp. 1). No knowledge of results was given, but *E* periodically indicated that *S*s were "doing fine." A 10¢ reward item was provided at the end of the session.

Results

On each trial, *S*s were required to respond either YES if they had seen the central character before or NO if they had not; if they responded YES they were further required to class it as either identical or similar to the previous occurrence of that character. These two aspects of the response will be treated separately.

Regarding the initial YES–NO decision, a correct response for either an identical or similar item was YES, for a new item, NO. Figure 1 shows the proportion of correct responses as a function of item type and retention interval. The probability of correctly identifying a new item remains consistently high, while the probability of correctly identifying either an identical or similar item as old drops with time.

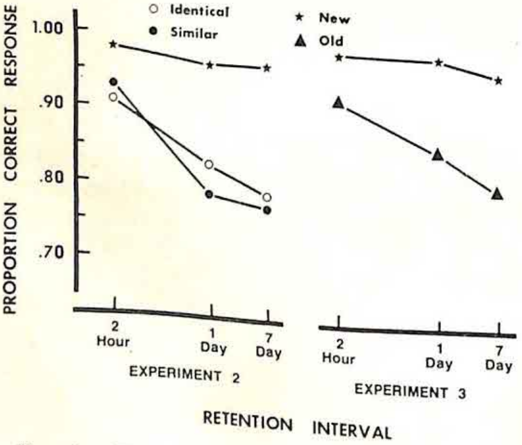


FIG. 1. The proportion of correct responses on the initial YES–NO (OLD–NEW) decision as a function of item type and retention interval.

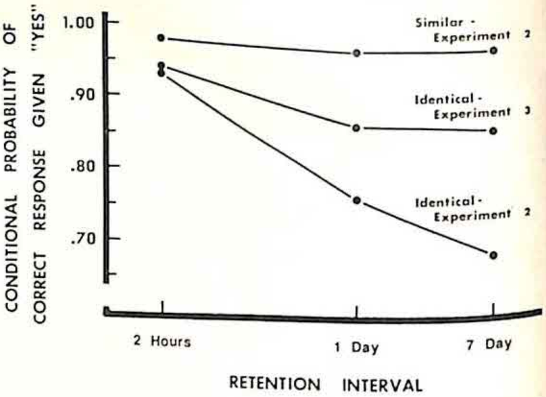


FIG. 2. The conditional probabilities of correctly identifying identical and similar items given that the item had previously been identified as old.

A 3 (Retention Interval) × 2 (Item Type) mixed analysis of variance was conducted on the errors on the old items only. Performance decreased significantly as a function of retention interval, F (2, 15) $= 21.00, p < .001$, but there was no reliable difference as a function of item type.

Given that *S* responded YES to an old item, he was further asked if it was identical or similar to a previously seen item. Figure 2 shows the conditional probability that this response was correct, given that *S* responded YES, as a function of item type and retention interval. The probability of responding SIMILAR to similar items remains high across retention intervals, but the probability of responding IDENTICAL to identical items drops sharply with time. The reliability of these effects was assessed by a 3 (Retention Interval) × 2 (Item Type) mixed analysis of variance. Confirming the visual impression, the Item Type × Retention Interval interaction was reliable, F (2, 15) $= 4.82, p < .025$. The fact that performance was relatively poor with the identical items at longer lags also led to reliable main effects due to item type, F (1, 15) $= 33.09, p < .001$, and retention interval, F (2, 15) $= 5.22, p < .025$.

Contingent performance measures.—The data which have been presented thus far have been the a priori data, the proportions of correct responses as a function of type of test item. Thus, $p(R_s|S_s)$ may be read as the probability that *S* responded SIMILAR given that the stimulus was in fact similar, and $p(R_i|S_i)$ the probability that *S* responded IDENTICAL, given that the stimulus was identical. Thus, $p(R_s|S_s)$ remains consistently high across retention intervals, whereas $p(R_i|S_i)$ declines with retention interval. One additional way of looking at the data is to investigate the a posteriori measures, the accuracy rates as a function of *S*'s response. If this is done, a clear trend emerges. The probabilities of *S*'s being correct, given that he responded IDENTICAL, i.e., $p(S_i|R_i)$, were .97, .95, and .94 at retention intervals of 2 hr., 1 day, and 7 days, respectively. The corresponding figures for $p(S_s|R_s)$ are .92, .80, and .75.

A similar pattern of results was obtained on the original YES–NO decision. Letting "o" and "n" stand for old and new, respectively, again the major finding is that $p(R_n|S_n)$ remains constant across retention intervals, whereas $p(R_o|S_o)$ declines with retention interval. Again, however, the $p(S_o|R_o)$ remains consistently high, with values of .97, .95, and .95 at retention intervals of 2 hr., 1 day, and 7 days, respectively. The $p(S_n|R_n)$ values decline with retention interval, the figures being .92, .83, and .75.

Discussion

The probabilities of recognizing old and new items can be compared with previous data (Brown & Scott, 1971) obtained under comparable conditions. At the 7-day retention interval, correct recognition of old items was 77% (averaged across identical and similar items), compared with 78% correct recognition of comparable once-seen items reported previously. Similarly, the probability of correctly identifying a new item was the same for both studies (.96). Thus, the addition, in this study, of both identical and similar test items, did not influence the ability to recognize recurring characters.

When the conditional probability of correctly classifying an old character as either identical or similar is considered, a puzzling pattern of results emerges. The probability of identifying a similar item (as similar) remained above .95 across retention intervals, but the probability of correctly identifying an identical item (as identical) fell to .68 at the longest interval. While it is possible that with longer retention intervals *S*s are less able to tell identical from similar items, this would not lead to the large difference between identical and similar items. If *S*s were simply less able to make the discrimination, performance would be impaired on both types of items.

One possible explanation of this difference lies in terms of a response bias. At all retention intervals, *S*s made more SIMILAR than IDENTICAL responses, with this difference increasing with retention interval. The percentages of similar responses were .52, .60, and .64 at retention intervals of 2 hr., 1 day, and 7 days, respectively. As the retention interval increased, details of the stored image would presumably fade. In time, all recurring items, whether identical or similar, would appear only similar to the remains of the original trace. Consequently, *S*s would have to adopt some criterion of similarity against which to judge which items are identical (very similar) and which items are similar (less similar) to the original trace. It appears that these *S*s adopt a very conservative criterion for terming an item identical. In essence, if they are sure the item is identical, they term it IDENTICAL; if they are unsure, they respond SIMILIAR. At longer retention intervals, details of the stored image have presumably faded, and fewer of the identical test items meet this conservative criterion, with the consequence that the number of similar responses increases. Some evidence consistent with this interpretation comes from a consideration of the contingent response measures. The probability that an item was in fact identical, given that *S* responded IDENTICAL, was consistently around .95, whereas the probability that the item was similar, given that *S* responded SIMILAR, dropped to .75 at the 7-day retention interval. That is, while *S*s are on occasion wrong when they respond SIMILAR, they are almost never wrong when they respond IDENTICAL. Again, this suggests that *S*s maintain a very conservative criterion for terming items IDENTICAL.

More relevant to the present question, the operation of such factors may make it impossible to use the continuous recognition memory procedure to obtain an accurate (criterion-free) estimate of *S*s' ability to distinguish similar from identical items. One alternative would be to use a forced-choice procedure as a way to minimize the criterion difficulties.

Experiment III

In Exp. III, the basic design of Exp. II was repeated using a modified forced-choice procedure which allowed the inclusion of the major features of the Exp. II procedure. On each test trial, two pictures of the same character in different poses were presented, and *S*s were required to tell if they had previously seen that character. Thus, *S*s were still required to indicate old and new items by the continuous recognition procedure. If *S*s said they had seen the character previously, they were further required to point to the specific picture they had seen before.

Thus, Exp. III was conducted in part to extend the results of Exp. II and to obtain an additional estimate of *S*s' ability to distinguish similar and identical items. Further, if the results obtained in Exp. II do reflect the operation of a response bias, it would be expected that performance on the forced-choice tests would lie between performance on the identical and similar items in Exp. II. The response bias would have tended to inflate performance on the similar items and depress performance on the identical items. The elimination of the response bias through the use of a forced-choice procedure should result in performance level midway between those obtained in Exp. II.

Method

Subjects.—A new sample of 20 preschool children were tested, and 18 *S*s, 10 boys and 8 girls, passed both pretraining and the immediate retention test. They ranged in age from 41 to 65 mo., with a mean of 57.

Stimulus materials.—The stimuli again consisted of pictures of people and animals cut from children's books. There were 80 triple sets (AAA′, consisting of two identical poses of some character and one similar pose of the same character) and 80 double sets (BB′, consisting of two similar poses of the same character). The inspection series consisted of 80 cards, each containing one picture pasted in the center. These single pictures were the first A members of the triple sets. The pretraining and retention stacks consisted of cards which contained two pictures. Each card contained one character in two different poses. For the old items, the second and third members of the triple sets (A and A′) were used, so that one pose was identical to that in the inspection stack and the remaining pose was similar. The new cards contained the double sets (BB′), consisting of two poses of a character not seen in the inspection set. Thus, the AA′ cards contained old characters and the BB′ cards contained new characters.

Procedure.—The experiment was conducted in two stages. The *S*s were first shown the inspection series consisting of 80 single pictures (exactly the same as in Exp. II). These consisted of the to-be-tested items which were one of the identical members (A) of the triple sets.

For the pretraining, immediate test, and retention test the same kind of items were presented. There were two types of items, old and new. Both old and new items contained two pictures consisting of two different poses of the same character. On old items (AA′), one of the poses (A) was identical to a picture that had occurred in the inspection set; the other picture (A′) was of the same character in a similar pose. On the new items (BB′), there were two similar poses of a character that had not occurred before in the inspection series. As each card was presented *S*s were asked, "Have you seen him(her) before?" If they responded positively, they were required to point to the one exactly the same as the picture they had seen before in the inspection series. They were periodically reminded that the picture must be exactly the same: "He must be doing the same thing."

During pretraining a stack of 20 cards was presented, 10 old cards and 10 new. The *S*s were trained to respond YES to characters seen in the inspection series and NO to new characters and, following a YES response, trained to point to the identical pose. Following pretraining, *S*s were given an immediate retention test consisting of 10 old and 10 new cards and were required to reach an 80% criterion of correctly identifying new from old items, and 80% correct on pointing to the identical picture.

The *S*s who passed both pretraining and the immediate test criteria were randomly assigned to one of three retention intervals of 2 hr., 1 day, or 7 days. At the designated interval, *S*s were reshown the pretraining pack and reminded of the game. Following this, they were presented a stack containing 120 cards, 60 old (AA′) and 60 new (BB′) items. No knowledge of results was given. A 10¢ toy was provided as a reward following both testing sessions.

Results

On each test trial, *S*s were shown two similar poses of the same central character. They were first required to indicate whether or not they had seen that character before. The percentage of correct responses on old and new items as a function of retention interval is shown in Fig. 1. As can be seen the probability of correctly identifying new item remains high across retention intervals. However, there is a reliable decrease in performance on old items as a function of retention interval, $F\ (2, 15) = 6.92, p < .01$.

Given that *S* correctly identified a character as old, he was further required to point to the pose identical to the one seen earlier. The conditional probabilities of a correct response, given a prior recognition that the character was old, are shown in Fig. 2 as a function of retention interval. The decrement in performance at long intervals is reliable, $F\ (2, 15) = 5.03, p < .025$. Also of interest, inspection of Fig. 2 reveals that the conditional probability of pointing to the identical item, under the forced-choice procedure of Exp. III, is approximately midway between the conditional probabilities of correctly identifying similar and identical items in Exp. II.

Discussion

Long-term recognition of old and new items was equally efficient in both long-term retention studies. The modified forced-choice method of Exp. III resulted in very similar YES–NO data to that produced by the traditional continuous recognition memory task. Considering performance following a retention interval of 7 days, in Exp. II the probability of identifying an old item was .77, averaged across similar and identical items, compared with .81 in Exp. III. Similarly, the probability of recognizing a new character was .96 in both studies.

The forced-choice procedure was used in an attempt to obtain a criterion free estimate of *Ss'* ability to discriminate identical from similar test items. The data indicated an accuracy level midway between that suggested for identical and similar items under the procedure of Exp. II. This finding is consistent with the suggestion that the declining accuracy for identical items in Exp. II was in part an artifact of the procedure used and reflected the operation of a conservative subjective criterion for assessing identity. When a forced choice between an identical and a similar pose was presented, *Ss* were able to identify the identical pose on at least 85% of all occassions.

GENERAL DISCUSSION

The data from these experiments provide information regarding two major points, the ability of young children to recognize minor changes in recurring pictures and the influence of subjective response criteria or biases. These two aspects of the data will be discussed separately.

The original question addressed by this series of experiments concerned the ability of young children to recognize minor changes in recurring pictures. The degree to which such changes could be recognized over time would indicate the amount of detail of the original picture which was stored and retained. The data from Exp. I indicated that considerable detail of the pictures was stored initially, with the data from Exp. III in particular indicating that these details faded slowly. Within the immediate tests of Exp. I and the longer retention intervals of Exp. III, identification of identical test items never fell below 85% correct. These data indicate that sufficient detail must be stored and retained to distinguish minor pictorial changes in recurring stimuli, at least when these changes involve the action of a central character. These data are contrary to the suggestion made by Reese (1970) that children's memory for visual information becomes "leveled," i.e., reduced to a figurative image stripped of idiosyncratic features.

In addition to the length of the retention intervals, Exp. I also differed from Exp. II and III in one other important respect. In Exp. I, *Ss* were alerted to the possibility that changes in the action of the central character would be important by the pretraining treatment and the use of the continuous recognition procedure. However, in Exp. II and III, *Ss* were not informed that the characters might be "doing something different" until after the first occurrence of all items (i.e., after the trace had originally been stored). That they could still identify action changes with considerable accuracy over long periods of time suggests either that all details of the picture were originally stored, or that the action of the character contributed largely to the meaning of the pictures and changes in meaning are noticed over long periods of time (Sachs, 1967). To separate these possibilities would require a comparison of different types of stimulus change involved in the similar test items. The character change could involve a different pose, one which implied an action change (running vs. eating) and another change which presented a different pose of the same action. If both types of similarity are noticed, then the image must contain considerable detail. If only action changes are noted, it would imply that the action of the character is a critical feature which remains as other less "meaningful" details fade.

The second major aspect of the results concerned the effects of response biases (or criterion effects) on preschoolers' recognition performance. The divergence in the similar and identical curves of Exp. II was attributed to a response bias effect. The *S*s were assumed to adopt a very conservative criterion for terming an item IDENTICAL. If they were uncertain, they responded SIMILAR. At longer retention intervals, as details faded, their uncertainty would increase, and the proportion of similar responses would also increase. The data from Exp. III, indicating a performance level midway between that on the similar and identical items from Exp. II, are consistent with such an explanation.

In general, the data from all the experiments are consistent in showing that the children adopt a very conservative criterion for saying either OLD or IDENTICAL. These two responses can be considered equivalent since, in both cases, *S*s are indicating that they have seen the picture before. When they say either NEW or SIMILAR, they are saying they have not seen that picture before. The same patterns of results were obtained in all phases of this research. As the retention interval increased, the probability of *S*s' responding NEW or SIMILAR increased, and the probability of correctly responding OLD or IDENTICAL decreased. However, given that *S*s responded either OLD or IDENTICAL, the probability that they were correct remained around .95. Our interpretation of this pattern of results is that *S*s' performance was affected by a bias to respond either NEW or SIMILAR whenever they were uncertain.

REFERENCES

Anisfeld, M. S., & Knapp, M. Association, synonymity, and directionality in false recognition. *Journal of Experimental Psychology*, 1965, **77**, 171–179.

Brown, A. L., & Scott, M. S. Recognition memory for pictures in preschool children. *Journal of Experimental Child Psychology*, 1971, **11**, 401–412.

Corsini, D. A., Jacobus, K. A., & Leonard, D. S. Recognition memory of preschool children for pictures and words. *Psychonomic Science*, 1969, **16**, 192–193.

Davis, G. A. Recognition memory for visually presented homophones. *Psychological Reports*, 1967, **20**, 227–233.

Hall, J. W., & Ware, W. B. Implicit associative responses and false recognition by young children. *Journal of Experimental Child Psychology*, 1968, **6**, 52–60.

Nickerson, R. S. Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology*, 1965, **19**, 155–160.

Nickerson, R. S. A note on long-term recognition memory for pictorial material. *Psychonomic Science*, 1968, **11**, 58.

Reese, H. W. Imagery and contextual meaning. *Psychological Bulletin*, 1970, **73**, 404–414.

Sachs, J. S. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, 1967, **9**, 437–442.

Shepard, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 156–163.

Underwood, B. J. False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, 1965, **70**, 122–129.

(Received November 1, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 63–65

THE ROLE OF LANGUAGE HABITS IN UNDERSTANDING LETTER SOUND SEQUENCES[1]

MICHAEL GODKEWITSCH[2]

University of Toronto

German-English bilinguals and unilingual English *Ss* heard German and English pentagrams in a random order. Letters were spelled sequentially in English, masked by a background white noise of either 74 or 83 db. Results supported the hypotheses that bilinguals would make more errors in English than native speakers, and that English unilinguals would err more on German than on English pentagrams. Under the high-noise condition this was even more pronounced: bilinguals did better than unilinguals did on German pentagrams. Aside from acoustic factors, language habits are important in processing letter sounds.

Miller and Isard (1963) showed the effects of linguistic familiarity upon perception: regular sentences were perceived more readily than syntactically or semantically anomalous ones. Following up on these observations, one may inquire whether other aspects of language similarly induce an implicit knowledge of rules whose effects are manifested in perception. In the following experiment, the perception of spoken sequences of letters was studied.

The transitional probabilities of sounds in a spoken language, and letters in a written language, are governed by phonological rules particular to that language. A bilingual whose first language is German and whose second is English is less familiar with the transitional probabilities of English than is a native speaker of English, independently of their familiarity with these sounds per se. Relative unfamiliarity must result in more errors in processing sequences of sounds bearing English characteristics. Further, when hearing conditions become difficult, and sequences of letters are presented following transitional rules of their first language, knowledge of the relevant transitional probabilities must put bilinguals at an advantage over unilingual English speakers.

METHOD

Design.—Eighty pentagrams from either English or German were presented in one randomly determined order to 20 English unilinguals and to 20 German-English bilinguals. The *Ss* were randomly assigned to one of two conditions: a background white noise of 74 or 83 db. Thus, a 2 × 2 × 2 × 2 factorial design was obtained with repeated measures on two factors: type of item and level of background noise.

Material.—Pentagrams, combinations of five letters, were chosen as stimulus material in order to establish some homogeneity in the stimuli and to equate them for meaning for *S* groups. The pentagrams were spelled out on tape, using English spelling (Z-zed). The pentagram ICHLY, for instance, would sound like, eye—see—aitch—el—why. The first six items were for practice, and the other 80 formed the test series. Half of the experimental pentagrams were English, taken from Mayzner, Tresselt, and Wolin (1965). The other half were German, selected from Schönpflug (1970). The two groups were matched for frequency in the particular language and for distribution of letters. Criteria for selection of pentagrams were: (*a*) no two consecutive identical letters, (*b*) no English pentagram should also exist in German and vice versa, (*c*) no pentagram should form a word, (*d*) well-known acronyms (JVLVB, USAFB etc.) were avoided.

Presentation.—The pentagrams were recorded by a tenor voice and presented at a rate of one per 5 sec., each pentagram being of approximately 1.5 sec. in duration. The *Ss* listened through earphones, and after each pentagram wrote the five letters on an answer sheet. The *Ss* were instructed to guess if necessary. The loudness of the spoken letters was about 91 db. A masking white noise of 74 db. was added for one half of the *Ss* in the uni- and bilingual groups, and 83 db. of white noise for the other half.

Subjects.—The unilinguals were male and female undergraduate and graduate students at the University of Toronto. The bilinguals were mainly students of about the same ages, in disciplines other than psychology. Since the results for female *Ss* differed in no way from those of males, the data were pooled.

---

[1] The research was supported by the Department of Psychology of the Unniversity of Toronto.

[2] Requests for reprints should be sent to Michael Godkewitsch, Department of Psychology, University of Toronto, Toronto 181, Ontario, Canada.

RESULTS

The scoring criteria were as follows: each processed letter differing from the recorded letter counted as one error, except when two adjacent letters were switched. In the latter case one error was counted. When three adjacent letters were mixed in order, two errors were counted. Any omitted letter or an incorrectly added letter counted as one error.

Analysis of variance showed three significant effects: the level of background noise, $F\ (1,\ 36) = 43.62$, $p < .001$, the two-way interaction between *S*s' first language and the language from which the pentagrams were taken, $F\ (1,\ 36) = 62.96$, $p < .001$, and the triple interaction of these three factors, $F\ (1,\ 36) = 6.13$, $p < .05$.

Table 1 shows the mean number of errors in the various treatment combinations. Analysis of the variation among groups confirmed both main predictions: bilinguals made more errors in English pentagrams than unilingual English speakers, $p < .005$. Also, bilinguals did significantly better on German than on English pentagrams, $p < .005$. Under the weaker noise condition bilinguals as a group made significantly more errors than unilinguals did, $p < .01$, probably due to familiarity with English sounds. For German pentagrams only there was no significant difference. Under the stronger noise condition unilinguals erred more in processing German pentagrams than did bilinguals, $p < .01$.

TABLE 1

MEAN NUMBER OF INCORRECTLY PROCESSED LETTER SOUNDS

Stimuli	Level of background white noise		
	74 db.	83 db.	$\bar{X}$
English pentagrams			
Bilinguals	11.85	24.80	18.33
Unlinguals	4.10	18.45	11.28
German pentagrams			
Bilinguals	8.75	18.60	13.68
Unilinguals	7.45	24.55	16.00
$\bar{X}$	8.04	21.60	

DISCUSSION

Letter sounds are probably guessed on the basis of two sources: the incomplete, masked or ambiguous information from the particular stimulus source, and the already processed letter sounds in the particular sequence. The *S* picks his missing letter from a number of alternatives on the basis of the phonological grammatical restraints of his first language, especially under difficult hearing conditions. A more parsimonious explanation of why bilinguals make more errors in English pentagrams than unilinguals do, namely that bilinguals are simply less familiar with English sounds, must be rejected: all sounds in this experiment were English pronounced. There would therefore be no reason for unilinguals to err more when processing German sequences, unless properties of the sequences per se are of importance. Also, if the simpler explanation were correct, there should have been a significant main effect for the factor representing *S*s' first language. Error counts, furthermore, indicate that the error data were not based on only a few sounds consistently misinterpreted. If this had been the case the phenomenon would have been weak indeed.

Of course some sounds do stand up to distortion better than others and account for more errors. However, if transitional probabilities between sounds determine processing of ambiguous sounds, then one would expect distortions of German sequences in the direction of English by unilinguals, and of English sequences toward German by bilinguals. To test this hypothesis a list of the 126 reproduced letter sequences containing more than three errors was presented to a panel of four bilinguals. They were asked to reach an agreement about the categorization of the "pseudo-words as English, ambiguous, or German. Of the 7 sequences reproduced by the bilingual experimental *S*s, 65 were categorized as German, 15 a ambiguous, and 3 as English. Of the 53 sequences reproduced by unilinguals, 32 were categorized as English, 7 as ambiguous, an 14 as German. Errors, made by bilinguals, a preponderantly toward their first languag $\chi^2(2) = 91.02$, $p < .01$, where unilinguals ten to distort ambiguous sounds toward the own language, $\chi^2(2) = 18.79$, $p < .01$. Th finding, together with the fact that the erro for single letters were equally distributed ov the five positions in the sequences indicat that the errors reflected difficulties in unde standing as well as in remembering. Kints

(1970, p. 191) has stated that acoustic confusions override language factors in relative importance in short-term memory. However, despite the use of sequential auditory presentation which favors acoustic coding (Baddeley, 1971), it was found that to the extent that short-term memory is operative in the task, language habits, next to acoustic factors, codetermine the processing of spoken sounds.

Finally, the manner of presentation (English spelling) induced a set toward perceiving English sequences of sounds. The *S*s did not know that their linguistic background was related to their participation in the experiment, nor did bilinguals recognize German sequences as such. The data of the unilingual *S*s support previous findings (cf. Miller & Isard, 1963) that, because grammar constrains word order, sequences conforming to *S*s' language habits are more easily encoded and recalled than are sequences that do not. The data of bilinguals show that, in addition, language habits which are not relevant to the task at hand can "supplement" relevant habits when the stimulus situation becomes more ambiguous.

REFERENCES

BADDELEY, A. D. Language habits, acoustic confusability, and immediate memory for redundant letter sequences. *Psychonomic Science*, 1971, **22**, 120–121.

KINTSCH, W. *Learning, memory and conceptual processes.* New York: Wiley, 1970.

MAYZNER, M. S., TRESSELT, M. E., & WOLIN, B. R. Tables of pentagram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements*, 1965, **1**, 145–185.

MILLER, G. A., & ISARD, S. Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 217–228.

SCHÖNPFLUG, W. n-Gramm-Haüfigkeiten in der Deutschen Sprache. IV. Ausgewählte Pentagramme. *Zeitschrift für Angewandte und Experimentelle Psychologie*, 1970, **17**, 139–157.

(Received November 3, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 66–71

INCREASING ITEM ACCESSIBILITY IN FREE RECALL[1]

MARILYN A. BORGES[2]

University of California, San Diego

Previous free recall studies have suggested that *S*s have more items available in storage than are accessible on any one recall trial. To decrease these retrieval or recall limitations, two groups of *S*s in Exp. I were presented a categorized word list and given two successive recall trials. Only Group 2 was given "multiple-repetition" instructions prior to their Second Recall. On their Second Recall, Group 2 recalled more list items, gave more extralist intrusions, item repetitions, and took longer to recall. It is suggested that motivational influences of the multiple-repetition instructions affected *S*s' recall-time termination rules, and this contributed significantly to their increased recall performance. Other variables of item repetition and intrusions had no such obvious influence. Experiment II confirmed that motivational incentives alone increase both recall time and performance.

Previous free-recall studies by Tulving (1967), Robinson and Kulp (1970), and Tulving and Pearlstone (1966) have suggested that *S*s have more items available in storage than are accessible on any one recall trial. If after having been presented a list of unrelated words, *S*s are given several successive recall trials with no additional learning trials, *S*s will on each successive recall trial recall a few additional "new" items from the list, but also "forget" a few of the previously recalled items.

Keppel and Mallory (1969) attempted to discover whether *S*s were withholding some of their responses for fear of making errors. Giving *S*s instructions to either "guess freely" or "make very few errors" in recalling a list of unrelated words, Keppel and Mallory did succeed in changing *S*s' error frequencies, but there was no difference in the number of items correctly recalled in either of these two groups as compared with a group of *S*s who were given standard free-recall instructions. Similarly, Cofer (1967, Exp. 1 and 2) found that when *S*s were forced to give responses that equaled the number of items presented, the number of correct responses did not appreciably increase.

In another attempt to free *S*s from any self-imposed retrieval limitations, or "selector mechanisms," Bousfield and Rosner (1970), in a multitrial free-recall experiment, compared the recall results of a group of *S*s who received "uninhibited recall instructions"; i.e., *S*s were told to emit all words occurring to them during recall, with a group of *S*s who received standard free-recall instructions. They found that recall was similar for the two groups on the first study-recall trial, although the "uninhibited" group had significantly higher recall than the "standard" group on the last, or fifth, recall trial. Bousfield and Rosner concluded that the facilitating effects of the uninhibited recall instructions were due to less intertrial forgetting of previously recalled items and were independent of the amount of overt errors produced.

In the experiments cited above, lists of unrelated words were used. Cofer (1967, Exp. IV) presented categorized lists of either high- or low-taxonomic-frequency nouns to *S*s and again forced them to provide a response for every item presented during study. Whereas the forced recall results showed little improvement for the low-taxonomic-frequency nouns, the forced recall did measurably enhance recall for the high-taxonomic-frequency nouns. However, since "guessed" responses under such conditions are typically high-taxonomic-frequency items from the presented categories, it may be that the increased recall for the high-frequency lists was due to guessing alone.

---

[1] This research was supported by National Science Foundation Grant GB 20798 and United States Public Health Service Grant MH 15828 from the National Institute of Mental Health. The author wishes to thank George Mandler and Richard Meltzer for their invaluable assistance.

[2] Requests for reprints should be sent to Marilyn A. Borges, Department of Psychology, University of California, San Diego, P. O. Box 109, La Jolla, California 92037.

Thus, the experimental literature to date suggests that *S*s will recall a relatively constant number of responses over successive recall trials, although there are fluctuations of various items between recall trials. No experimental attempts to decrease *S*'s "retrieval limitation," i.e., to increase the size of the response set and include more stored information on any one recall trial, have been clearly successful.

The purpose of the present study is to explore some of the variables which may affect *S*s' retrieval strategies in a free-recall paradigm and, if possible, to decrease the degree of retrieval or recall limitations.

EXPERIMENT I

Method

Design.—Two groups of *S*s were presented a list of categorized words for two study trials followed by two successive category-cued recall trials. All *S*s received "standard free-recall" instructions, prior to their first recall, with half of the *S*s hearing again the standard free-recall instructions (Group SI) and half "multiple-repetition" instructions (Group MRI) prior to their second recalls. A third group of *S*s constituted a control group (CG) needed to assess the effects of guessing on recall performance.

Subjects.—The *S*s were 36 undergraduates fulfilling a course requirement. Twelve *S*s were randomly assigned to each of the three groups.

Lists.—Each *S* in Groups SI and MRI was shown one list of 60 words. The list consisted of 10 words from each of six semantic categories. The words from each category ranked from 6 through 15 in their frequencies of occurrence as responses to the category name in the norms of Loess, Brown, and Campbell (1969). The six categories used were vegetables, articles of clothing, parts of the body, parts of a building, occupations or professions, and natural earth formations. Four different orderings of the list were constructed, with word sequence randomized and with the restriction that no two words from the same category be adjacent.

Procedure for Groups SI and MRI.—All *S*s learned a 60-word categorized list prior to this experiment in order to familiarize them with the test procedure. Instructions stated that the experiment was concerned with memorization of lists. The *S*s were fully informed of the nature of the lists and were given a typed card containing a randomized listing of the six category names. The *S*s were told they could use the card during recall, and the card remained in front of *S*s during list presentation and subsequent recall trials. The words were presented on a Stowe memory drum at a 2-sec. rate with *S*s saying each word aloud during presentation. Each list was presented in two of the four random orders with the recall tests given after the second presentation of the list.

For the First Recall, all *S*s were instructed to recall orally as many of the words from the list as they could remember. Recall was terminated when *S* indicated to the *E* that he could not recall any more words.

For the Second Recall, *S*s in Group SI were asked to try again to recall all the words from the list (same instructions as given previously). The *S*s in Group MRI, in addition to the previous recall instructions, were given the following multiple-repetition instructions: "Don't worry about saying the same words over and over . . . just keep going over the list words that you know while searching for new words." In addition, during the Second Recall period, *S*s in Group MRI were asked to write down those recalled words which they knew had not been recalled during their previous recall period. If they were not certain whether they had recalled the word or not last time, they were to write the word and then place a question mark after the word. Again, recall terminated when *S* indicated to *E* that he could not recall any more words.

Procedure for Group CG.—Group CG *S*s were tested subsequent to Groups SI and MRI. Each *S* was given a sheet of paper containing a list of the category names and the items emitted during the Second Recall task by one *S* from Group SI. Each *S* was told that he had a list of all the words that a previous *S* had recalled from a presented list of 60 words, 10 words from each of the listed categories. The *S*s were told that the purpose of this experiment was to see whether *S*s in such a recall task tended to "guess" additional items when they could remember no more list words. Hence, CG *S*s' tasks were to fill in all blanks with their "best guesses" of other items which may have appeared on the original 60-item list.

From zero to two blanks appeared at the end of each category listing for each *S*. The total number of blanks per category was the arithmetic difference between the total items per category produced by all Group MRI *S*s and the total items per category produced by all Group SI *S*s in the Second Recall tasks. The resulting total number of blanks per category were then distributed as evenly as possible across the categorized lists given to Group CG *S*s.

Results and Discussion

Recall.—Figure 1 shows the mean recall data for *S*s in Groups SI and MRI during their First and Second Recall tasks. Analyses of variance on the recall, intrusion, and repetition data showed that the interactions between groups and recalls were significant for all three measures, $F\ (1,\ 22) = 6.29$, 6.56, 7.76, $p < .05$, respectively.

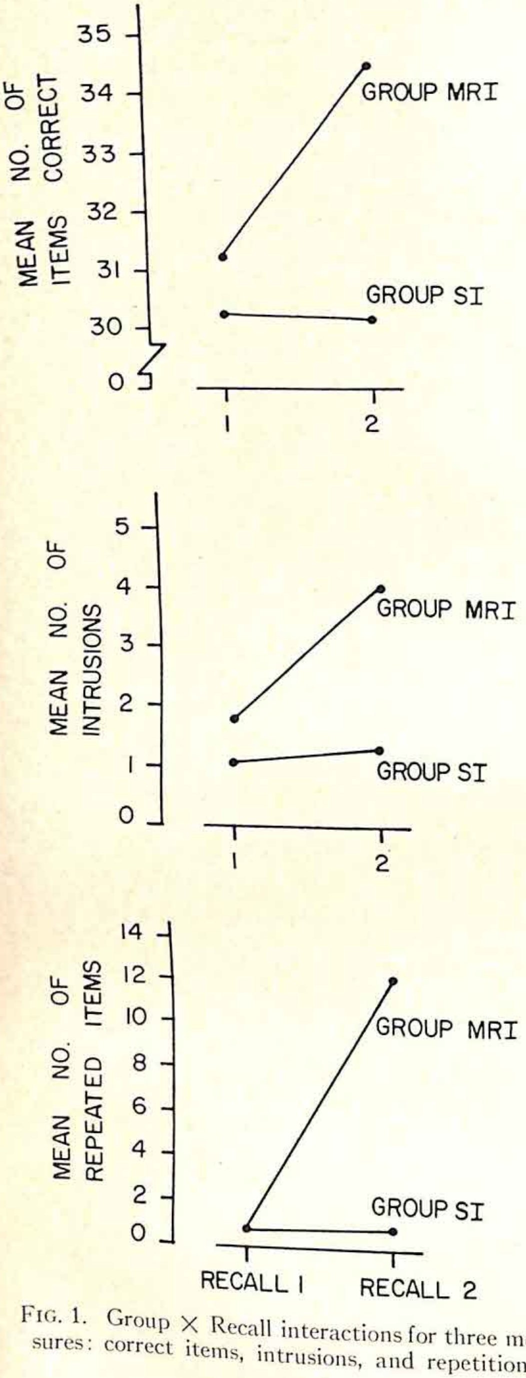


FIG. 1. Group × Recall interactions for three measures: correct items, intrusions, and repetitions.

The *S*s in Group MRI recalled more list words in their Second Recall than did *S*s in Group SI, but MRI *S*s also recalled a higher number of intrusions, thus increasing their total number of items produced (ignoring repetitions) over that of SI *S*s by 84 items (52 new words, 32 intrusions). Of the 84 items guessed by CG *S*s, only 4 appeared on the original presentation list. Thus, it is unlikely that guessing had any significant effect on increasing the Second Recall performance of Group MRI over that of Group SI.

The *S*s in all conditions determined their own recall times. There was no difference in total time between *S*s in Groups SI and MRI on their First Recalls (median time = 2 min., 41 sec.), but *S*s in Group SI took a median time of 2 min. on Recall 2, while *S*s in Group MRI took 5 min., 19 sec. on Recall 2.

Table 1 shows that Group MRI's increased Second Recall performance is made up of two separate components: (*a*) Group MRI *S*s forgot significantly fewer words that they had previously recalled correctly on their First Recall task than Group SI *S*s, and (*b*) Group MRI *S*s recalled significantly more new words during their Second Recall that they had not recalled during their First Recall than did Group SI *S*s, $t(22) = 2.18$, $p < .05$ and $t(22) = 1.89$, $p < .05$, respectively.

All *S*s in both groups recalled by category; i.e., they recalled a cluster of words from one category, then a cluster of words from another category, etc., and then would sometimes recall a few additional words from categories entered earlier. To analyze output-position effects, the output order of recalled words was separated into four parts across categories. The initial category cluster (ICC), the first time *S* recalled any words from a category, divided into thirds, comprised three of the parts. The fourth part was words recalled from a category *after S* had completed his ICC for that category; i.e., *S* recalled additional words from a previously entered category, having recalled other-category items in the interim.

As indicated in Table 2, list words recalled by *S*s in Group SI on Recall 1 only ("forgotten" on Recall 2) tend to be words that were output in the first-third ICC, 19 out of 42 words. This contradicts a strength theory which predicts that the first items recalled have the greatest memory strength or highest probability of being remembered, as well as the notion that

TABLE 1

Mean Recall Performance Analysis for Groups SI and MRI

Recall measure	Group SI	Group MRI
Words correct on Recall 1, but omitted on Recall 2	3.5	2.1
Words omitted on Recall 1, but correct on Recall 2	3.4	5.3

there may be certain key words which *S*s use to enter into categories during recall. It is very unlikely that this loss is due to any "echo box" transfer from short-term memory during Recall 1, since Group MRI did not show any difference in loss of items across the three thirds of their ICCs. The new words that *S*s remembered in their Second Recalls had the highest probability of being recalled during the last third of the ICC, and for Group MRI *S*s also in subsequent recall. This indicates that as recall progresses within a category, the probability of new words becoming accessible increases.

Of the 64 list-words that *S*s in Group MRI recalled during their Second Recall that they had not recalled during their First Recall, *S*s recognized 81% of the words as "new," 5% as unsure whether new or old (previously recalled) and 14% as old. In addition, *S*s called 3% of the previously recalled words new and 3% were marked unsure whether old or new. These results agree with Robinson and Kulp's (1970) findings that *S*s' knowledge of the contents of a prior recall is quite accurate.

If encouragement of repetitions is the factor which significantly increased list-item recall, one would expect a high correlation over *S*s between number of words repeated and number of new words on the Second Recall. Similarly, one would expect a strong negative correlation between number of words repeated on Recall 2 and number of forgotten words from Recall 1. However, the product-moment correlations were $-.003$ and $-.23$, respectively, with .65 ($N = 12$) needed for significance at the .05 level. Likewise, if the increase in extra-list intrusions on Recall 2 significantly affected recall, then one might expect the number of new and the number of forgotten items to be correlated with number of intrusions. The correlations were .20 and .40, respectively, also not significant at the .05 level. These negative results are consistent with the findings of the previous studies cited earlier.

The *S*s in Group MRI took over twice as long for their Second Recall as did Group SI *S*s. The correlations between number of new items recalled on the Second Recall and actual recall times in the Second Recall, T_2, for the 24 *S*s in Groups SI and MRI is .43 ($p < .05$); between number of items recalled on Recall 1 but forgotten on Recall 2 and the Second Recall times is $-.39$ ($p < .10$). The two effects, combined, produce a correlation between total Recall 2 and Time 2 of .64. The correlation between total Recall 1 and the First Recall times was only .33; hence time was more related to recall performance on Recall 2 than on Re-

TABLE 2

Relative Output Positions of Once-Recalled List Items

Recall measure	Initial category cluster (in thirds)			Subsequent recall	Total N
	First	Second	Third		
No. of list-words per recall position on Recall 1; not recalled on Recall 2 (forgotten items)					
Group SI	19	8	11	4	42
Group MRI	7	6	8	4	25
No. of list-words per recall position on Recall 2; Not recalled on Recall 1 (new items)					
Group SI	5	7	22	7	41
Group MRI	4	7	29	24	64

call 1. Since some *S*s may consistently choose to take more time for recall than other *S*s, such individual differences may influence the number of new and forgotten words they recall, as well as their Second Recall times. Hence, partial correlations were computed using each *S*'s First Recall time, T_1, as a base line. The partial correlation coefficients between T_2 and number of new items recalled on the Second Recall, holding T_1 constant, is .50 ($p < .05$); between T_2 and number of items recalled on Recall 1 but forgotten on Recall 2, holding T_1 constant, is $-.56$ ($p < .01$). The correlation between T_2 and total Recall 2, holding T_1 constant is .63. Thus on the Second Recall task, those who spent more time retrieved both more new items *and* previously recalled items than *S*s who spent less time.

It is suggested that the increase in recall performance due to longer *S*-determined Second Recall times may be related to the "demand characteristics" of the particular Second Recall instructions. Group MRI *S*s, as a group, may have been more highly motivated than the SI group due to the peculiarities of the MRI instructions; i.e., handing *S* a pencil and paper on which to write the new words may have been interpreted by *S* to mean that *E* expected that he would remember at least a few new items. However, each individual *S* may have been motivated differently, even within instructional conditions, and this internal motivation may have influenced his choice of "temporal termination rules" or limits that *S* places on the duration of his retrieval attempts. It may be that these *S*-determined termination rules constituted one crucial variable which affected recall performance.

Experiment II was designed to test whether or not motivational incentives alone could influence *S*-determined termination rules and, if *S*s did increase their recall times, whether the total number of items recalled also increased. Alternatively, perhaps just assigning *S*s a longer time to recall (*E*-determined termination rules) would produce an equivalent increase in recall.

EXPERIMENT II

Method

Design.—Four groups of *S*s were presented a list of categorized words for two study trials, followed by two successive category-cued recall trials for three of the groups, with one group having only one recall trial. The groups differed only in the recall instructions which were given immediately prior to each recall trial. The materials used were identical to those of Exp. I.

Subjects.—The *S*s were 48 undergraduates fulfilling a course requirement. Twelve *S*s were randomly assigned to each group.

Procedure.—All *S*s were tested under the identical conditions as *S*s in Groups SI and MRI of Exp. I, with the exception that recall instructions varied for each of the four groups. Group FF (Free, Free) *S*s were given the same standard recall instructions as Group SI from Exp. I (*S*s were *free* to determine their own recall termination times for both Recalls 1 and 2). The *S*s in Group FF were tested prior to the following groups. Group FC (Free, Constrained) *S*s were given the same recall instructions as Group FF with the addition of the following statement in their Second Recall instructions: "Keep trying to think of additional list items until I tell you to stop." The *S*s were then told to stop when *twice* the median time taken by Group FF *S*s on Second Recall had elapsed. (The Second Recall time was constrained).

Group C *S*s had only one recall trial. Their recall instructions also included the instructions to keep

TABLE 3

EXPERIMENT II: COMPARISONS OF MEAN PERFORMANCE

Recall measure	Group FF	Group FC	Group C	Group FM
Total items correct				
Recall 1	29.58	29.00	30.00	30.83
Recall 2	30.17	29.42		32.00
Recall times (Medians)				
Recall 1	2 min. 15 sec.	3 min. 17 sec.	4 min. 30 sec.[a]	2 min. 35 sec.
Recall 2	1 min. 55 sec.	3 min. 50 sec.[a]		3 min. 11 sec.
Intrusions				
Recall 1	.8	.6	.4	1.2
Recall 2	1.2	1.0		1.2

[a] *E*-constrained recall times.

recalling until *E* gave the signal to stop. The *S*s were told to stop at *twice* the median time taken by Group FF on first recall. Group FM *S*s were given the same recall instructions as Group FF with the addition of the following *motivating* statement included in their Second Recall instructions: "To provide you with some incentive, I will pay you 10 cents for every item you recall over 25. But to keep you from just guessing, for each item that you recall that was not on the presented list, I will deduct one item from your total recall score." The rationale behind the Recall 2 instructions for Group FM was that money would motivate *S*s to increase their recall scores with little specific influence on other factors such as item repetition or retrieval strategies which may have occurred in Exp. I due to the multiple-repetition instructions. By deducting intrusions from *S*s' paid recall scores, any increases in recall due to "pure guessing" should be minimized.

Results and Discussion

Recall times.—Table 3 shows the median *S*-determined or *E*-constrained recall times for the four groups. The motivational aspects of the Group FM Second Recall instructions did alter those *S*s' temporal termination rules; i.e., it lengthened the duration of their Second Recall periods, relative to Group FF.

Recall.—Table 3 shows the mean recall results for the four groups. An analysis of variance on the First and Second Recall data for Groups FF, FC, and FM showed no significant main effects, but subsequent Newman-Keuls tests on all possible pairs showed the mean recall performance of the FM group to be statistically greater than the other two groups, $p < .05$. An analysis of variance on the Recall 1 data for all four groups showed no significant differences between groups. Hence, *E*-constrained increases in either the First or Second Recall times had no significant effect on recall performance. This leads to the conclusion that recall time per se is not a crucial variable which affects recall performance, but that *S*'s own choice of recall termination rules may be.

It is suggested that the motivational effects of the promised money caused Group MR *S*s to increase the duration of their Second Recall, and thereby increased the total number of list items they recalled. Since the intrusion data showed no significant differences between groups, it is unlikely that the increase in MR's Second Recall is due to "pure guessing." Although Group MR's longer recall times and increased recall performances were not of the same magnitude as Group MRI's in Exp. I, it is likely that the peculiarities of the multiple-repetition instructions motivated *S* performance in somewhat the same way as the money motivated *S*s in Exp. II.

In conclusion, it is suggested that "motivational" instructional demands of verbal recall tasks can affect *S*s' performance by increasing the time *S*s *choose* to spend trying to recall. These *S*-determined termination rules constituted one factor that affected item recall. The other task variables of item repetition and extralist intrusions had no such obvious influence. This study indicates that there are more items stored in *S*s' memories which are accessible than are usually recalled in standard recall tasks.

REFERENCES

Bousfield, W. A., & Rosner, S. R. Free vs. uninhibited recall. *Psychonomic Science*, 1970, **20**, 75–76.

Cofer, C. N. Does conceptual organization influence the amount retained in immediate free recall? In B. Kleinmuntz (Ed.), *Concepts and the structure of memory*. New York: Wiley, 1967.

Keppel, G., & Mallory, W. Presentation rate and instructions to guess in free recall. *Journal of Experimental Psychology*, 1969, **79**, 269–275.

Loess, H., Brown, A., & Campbell, J. Cultural norms for items in 30 taxonomic categories. *Psychonomic Monograph Supplements*, 1969, **3** (7, Whole No. 39).

Robinson, J. A., & Kulp, R. A. Knowledge of prior recall. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 84–86.

Tulving, E. The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 175–184.

Tulving, E., & Pearlstone, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 381–391.

(Received November 8, 1971)

with a pause of approximately 45 sec. between lists. For this experimental phase, the timers were set at 1.3 sec. for the presentation of the 21 targets; otherwise, the procedure was identical to that of the practice trial.

RESULTS

Memory operating characteristics (MOCs) were determined with the purpose of assessing the degree to which the data conformed to the normality and homogeneity assumptions of signal detection theory. Hits ("old"/old) and false alarms ("old"/new), collapsed over Lists 1–3, were plotted on normal-normal coordinates. The two MOC curves showed minimal departure from linearity, indicating that the target and distractor distributions were reasonably Gaussian in form. The two MOC slopes, however, were less than unity, meaning that the target distribution was more variable than the distractor distribution for both conditions. This does not, however, appear to represent a serious violation of the homogeneity assumption, since the d′ and ds values were identical for both groups, ds being the appropriate index for data which are heterogeneous. The d′ measure was therefore employed as an index of sensitivity for all subsequent analyses. Differences between d′s were tested by the method of Gourevitch and Galanter (1967), while β differences were assessed by an extension of a technique developed by Smith (1969). The only signal detection analyses to be reported, however, will be those which augment the hit and false-alarm data.

TABLE 1

SUMMARY OF HIT AND FALSE-ALARM PROPORTION DATA

Dependent variable	Group	Lists			
		1	2	3	4
Hits	HH	.79	.79	.80	.82
	HL	.84	.84	.86	.91
	LL	.90	.86	.86	.86
	LH	.89	.89	.91	.85
False alarms	HH	.14	.14	.16	.18
	HL	.15	.18	.19	.10
	LL	.11	.12	.14	.14
	LH	.10	.10	.15	.17

The data of Groups H (HH and HL) and L (LL and LH) were examined over Lists 1–3 with the purpose of detecting any accumulation of interference over lists, the expectation being that this would be more pronounced for common than for rare words. Table 1 presents a breakdown of the hit and false-alarm data for all four groups for Lists 1–4. For hits, only the groups effect was statistically significant, F (1, 98) = 18.09, $p < .001$. For false alarms, both the groups, F (1, 98) = 4.81, and the lists, F (2, 196) = 3.45, both ps $< .05$, effects were significant but little or no interaction was evident. The *S*s of both groups, however, became less conservative over lists. The reaction time analysis involved another dimension in addition to those of groups and lists. This will be referred to as conditions, and it involved a comparison between hits and correct rejections. The basic unit of analysis was *S*'s median reaction time. The groups effect was statistically significant, F (1, 98) = 4.27, $p < .05$, the means of the medians being 1.60 sec. for Group L and 1.76 sec. for Group H. The conditions effect was also significant, F (1, 490) = 281.81, $p < .001$, the mean for hits being 1.48 sec., while that for correct rejections was 1.88 sec. Neither the lists effect nor any of the four interactions were statistically dependable.

It is clear from the preceding analyses that rare words were more readily detected than common words throughout the first three lists, this difference being apparent by the end of List 1. In order to determine whether this was also the case during the initial stages of the first list, the data were analyzed for the first 7 of the 42 test positions. This decision to divide the list into sixths was a purely arbitrary one. The proportion of hits for Group L was .905, while that for Group H was .804, the difference being statistically significant, z = 2.80, p = .003. These hit rates were comparable to those for the full list. The false-alarm data, however, exhibited a nonreliable reversal of this situation, the false-alarm probabilities being .038 and .035 for Groups L and H, respectively. The false-alarm rates for the early portion of List 1 were well below those for the full list. In order to determine whether this continued to be the case for Lists 2 and 3, a similar analysis was performed. The false-alarm probabilities on List 2 were .082 and .065 for Groups H and L, respectively, and these were not significantly different, $z = .65$, $p = .516$. The same situation prevailed for List 3, Group H = .089, Group L = .084, $z = .18$, $p = .857$.

It is evident from the preceding analyses that false alarms increased within lists, Groups H and L being comparable for the early portion of a list, with Group H exhibiting more false alarms by the end of the list. In order to achieve a more precise description of this phenomenon, which appeared to be stable over Lists 1–3, the data were collapsed over the three lists and false alarms were examined as a function of successive sixths of a list. A similar analysis was performed for each of the remaining four dependent variables. The hit and false-alarm data are presented in Fig. 1. The hit rates were stable over successive sixths, the $\chi^2(5)$ value for Group H being 5.56, $.30 < p < .50$, whereas that for Group L was 7.23(5), $.20 < p < .30$. False alarms, however, increased for both groups, $\chi^2(5)$ for Group H = 42.85, and $\chi^2(5)$ for Group L = 29.41, both ps $< .001$. The false-alarm rate was reliably higher for Group H than for Group L on the final sixth, $z = 3.69$, $p < .001$. Since the groups were comparable on the initial sixth, the slope of the false-alarm function was therefore steeper for Group H. Group H also exhibited a reliable β decrement from the first to the last sixth, whereas Group L did not. The reaction time analysis revealed the usual groups and conditions effects, as well as significant main effect for sixths, $F\ (5, 1078) = 13.28$, $p < .001$. The means of the median reaction times, in seconds, for the sixths were as follows, in succession: 1.59, 1.61, 1.71, 1.72, 1.75, and 1.82.

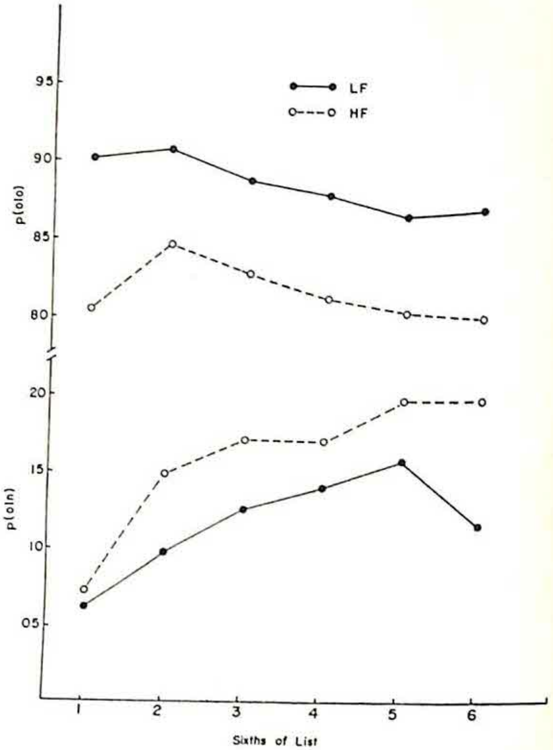


FIG. 1. Hit and false-alarm probabilities as a function of successive sixths of a list for the low- (LF) and high-frequency (HF) conditions.

The data for all four subgroups were examined for Lists 3 and 4 with the purpose of detecting release from interference, should it exist. The hit and false-alarm proportions are presented in Table 1. For the hit data, the groups effect was statistically significant, $F\ (3, 96) = 4.77$, $p < .005$, the overall hit rate for Group HH being lower than the other three groups. The only other significant effect for hits was the Groups × Lists interaction, $F\ (3,\ 96) = 3.29$, $p < .05$, this being primarily attributable to an increase in hits from Lists 3–4 on the part of Group HL, $t\ (96) = 1.78$, $p < .05$, as well as to a decrease in hits on the part of Group LH, $t\ (96) = 2.52$, $p < .01$. The absolute amount of change from Lists 3–4 did not differ reliably for these two groups, nor were there dependable shifts for Groups HH and LL. The false-alarm analysis also revealed a statistically significant Groups × Lists interaction, $F\ (3, 96) = 4.84$, $p < .005$, this being mainly attributable to a decrease in false alarms for Group HL, $t\ (96) = 3.65$, $p < .005$. The other three subgroups exhibited no reliable shifts in false-alarm rates. The only reliable β change was for Group LH, these *S*s becoming more conservative on List 4.

The hit and reaction time data for Groups

H and L were partitioned arbitrarily into seven blocks of three serial positions. Both sets of data were collapsed over Lists 1–3 since there was no evidence, in either instance, of a lists main effect or any interactions involving lists. The positions effect for hits was statistically significant, F (6, 588) = 7.49, $p < .001$, whereas the Groups × Positions interaction was not. The critical difference between pairs of means was .39 ($\alpha = .01$), Position 1 versus the other six positions satisfying this criterion, with all other comparisons failing to do so. The mean number of hits for Position 1 was 8.25, out of a total possible of 9, while the average of the remaining means was 7.57. The reaction time analysis, which was based on means of medians for hits only, revealed essentially the same phenomenon, the positions effect being statistically significant, F (6, 588) = 2.26, $p < .05$. The mean of the median reaction times for Position 1 was 1.43 sec., whereas the average value of the remaining six positions was 1.51 sec. Only Positions 1 versus 2 and 1 versus 6, however, exceeded the critical difference of .10 sec. ($\alpha = .05$).

The final analysis involved an examination of hit probabilities at Lags 0–6 versus 55–61, lag referring to the number of items intervening between the first and second presentation of a target on any given list. In order to achieve a sufficiently reliable analysis, the data were collapsed over Lists 1–3 for Groups H and L separately. There were no statistically significant differences between the short and long lag conditions for either group, a phenomenon previously observed by Shepard (1967).

DISCUSSION

The findings of the present investigation provide unequivocal evidence for the word-frequency effect in recognition memory, the superiority of rare over common items having become stable by the end of the initial list. The experimental hypotheses, however, received little support. Although the false-alarm data were consistent with the notion of accumulation of interference across lists, the phenomenon was no more pronounced for high-frequency than for low-frequency materials. Also, there was no evidence for release from interference as a result of *S*s being switched from one class of materials to the other. Release appeared to be, on the other hand, more related to the interlist interval than to a change in class. The findings, nevertheless, may still be rationalized within an interference framework. The hit and false-alarm data, when analyzed separately, appear to reveal evidence for two interference loci, one at storage and the other at the test phase of a list. Both sources evidently have a greater impact on common than on rare items. Previous investigators have been unsuccessful at uncovering these, since they have typically been concerned with either hits or false alarms, but not both except when calculating their joint effects and reporting these in terms of a single measure such as d′. The notion that interference is a retrieval phenomenon and hence should not be manifest in recognition memory (Bower & Bostrum, 1968; Wickelgren, 1967), therefore, will have to be carefully reexamined.

The hit rate for Group L was reliably higher than that for Group H by the end of the first sixth of the initial list. Beyond that point, hits for both groups had reached a steady state, there being no further change in hit rate either within a given list or from list to list. The two classes of items, therefore, were likely differentially interfered with during target storage. The changes in hit rates for Groups HL and LH on List 4 are consistent with this notion. The improvement on the part of Group HL may be attributed to a smaller amount of storage interference on List 4 in relation to that which accrued on earlier lists. Similarly, the lower hit rate on List 4 for Group LH presumably reflects more storage interference than that previously encountered. The primacy portion of the serial position function, which was observed for both hits and hit reaction times, suggests that storage interference appears to be, at least in part, proactive in nature.

The false-alarm data failed to reveal the frequency effect during the early stages of the initial list. Performance at that point was optimal for both classes of items and it was not until later on in the list that false alarms were reliably greater for common items. These data, therefore, provide strong support for the notion of test-phase interference whereby interference accumulates with increased exposure to distractors, the effect being most pronounced for high-frequency materials. This also appears to continue to develop from list to list, since false alarms exhibited a reliable increase over lists. The interlist interval, however, appears to allow for release from interference, since the false-alarm rate was low and was equivalent for Groups H and L during the first sixth of any subsequent list. The false-alarm data of Groups HL and LH are also consistent with the notion of differential amounts of test-phase interference for common and rare materials. The improvement in performance for Group HL may be attributed to release from interference as well as to a lower level of test-phase interference than that which had developed on previous lists. The absence of any reliable performance change for Group LH was likely due to two opposing tendencies of approximately equal magnitude, one being greater test-phase interference on List 4 than on previous lists, and the other presumably being release from interference in some combination with whatever was responsible for these *S*s becoming more conservative on List 4.

A number of other phenomena are of interest, although none has direct implications with respect to the notion of two interference loci. The longer reaction times for correct rejections than for hits is incompatible with a notion of exhaustive serial search but not necessarily with one of self-terminating serial search. These data may be embarrassing for those who claim that recognition memory tasks minimize retrieval demands. For a more detailed account of this issue, see McCormack (1972). The absence of a reliable recency effect may be attributed to the backward-counting task, which might be expected to interfere with the transfer of late items from a short-term to a long-term store. This interpolated task was also likely responsible for the failure to demonstrate any dependable lag effect, a zero-lag item having an actual lag of 75 sec. The trend toward less conservative behavior over Lists 1–3 for *S*s of both groups was likely a reaction to an increasing inability to discriminate between targets and distractors. This phenomenon of decreasing β values was also observed within lists, but only for *S*s of Group H. This finding is not particularly surprising since these *S*s were those most affected by test-phase interference.

REFERENCES

Bower, G. H., & Bostrum, A. Absence of within-list PI and RI in short-term recognition memory. *Psychonomic Science*, 1968, **10**, 211–212.

Ekstrand, B. R., Wallace, W. P., & Underwood, B. J. A frequency theory of verbal discrimination learning. *Psychological Review*, 1966, **73**, 566–578.

Garton, R. F., & Allen, L. R. Familiarity and word recognition. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 385–389.

Gorman, A. M. Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology*, 1961, **61**, 23–29.

Gourevitch, V., & Galanter, E. A significance test for one parameter isosensitivity functions. *Psychometrika*, 1967, **32**, 25–33.

McCormack, P. D. Recognition memory: How complex a retrieval system? *Canadian Journal of Psychology*, 1972, **26**, 19–41.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 1968, **76** (1, Pt. 2).

Paivio, A., Yuille, J. C., & Smythe, P. C. Stimulus and response abstractness, imagery, and meaningfulness, and reported mediators in paired-associate learning. *Canadian Journal of Psychology*, 1966, **20**, 362–377.

Shulman, A. I. Word length and rarity in recognition memory. *Psychonomic Science*, 1967, **9**, 211–212.

Schulman, A. I., & Lovelace, E. A. Recognition memory for words presented at a slow or rapid rate. *Psychonomic Science*, 1970, **21**, 99–100.

Shepard, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behaviour*, 1967, **6**, 156–163.

Smith, J. E. K. *A note on the standard error of the estimates of TSD parameters from "yes-no" experiments.* (Tech. Rep. No. 69–8) Ann Arbor: Mathematical Psychology Program, University of Michigan, 1969.

Sumby, W. H. Word frequency and the serial position effect. *Journal of Verbal Learning and Verbal Behaviour*, 1963, **1**, 443–450.

Wickelgren. W. A. Exponential decay and independence from irrelevant associations in short-term recognition memory for serial order. *Journal of Experimental Psychology*, 1967, **73**, 165–171.

Wickens, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

(Received November 11, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 78–89

MEMORY PROBES DURING TWO-CHOICE, DIFFERENTIAL REWARD PROBLEMS [1]

GORDON A. ALLEN [2]

Miami University

In a binary choice situation, information concerning the rewards associated with the alternatives was given to *S* using either the full information (FIN) procedure (reward value of both alternatives made known) or the partial information (PIN) procedure (reward value of chosen alternative made known). In Exp. I, *S* was required to estimate the reward associated with each stimulus after varying amounts of training on the choice problems. Most of the *S*s in PIN conditions learned one of the reward values, but there was evidence that this occurred after *S* made his last choice error. In FIN conditions, *S* did not typically learn either reward value. In Exp. II, after varying amounts of training, *S*s were tested on a set of problems constructed by repairing training stimuli. Afterwards, *S* also gave estimates of each reward value. The results of FIN conditions supported the hypothesis that *S*s learn, in an all-or-none fashion, which alternative is correct, but do not learn the value of a particular reward. The results of PIN conditions supported the hypothesis that *S*s acquire partial information about each reward, but do not learn the explicit value; rather, *S* encodes each reward, the code being a function of the absolute value of the reward.

In a binary choice experiment with differential rewards, *S* is presented with a list of problems; each problem presents a binary choice with rewards associated with each alternative. If *S* is given information about the reward values using the full information (FIN) procedure, *S* is shown the reward value he receives for the choice he made and also the reward value of the alternative he did not choose. Using the partial information (PIN) procedure, he is shown only the reward value of the alternative he chose.

Keller, Cole, Burke, and Estes (1965) conducted an experiment which contrasted the two information conditions for various combinations of reward values. They found that the rate of learning was directly related to the difference in reward value under the PIN procedure, but that the differences disappeared using the FIN procedure. They captured many aspects of choice learning using a model which assumed that the rate of forming associations between stimulus alternatives and rewards is independent of reward magnitude. It was assumed that the explicit value of each reward is learned in an all-or-none fashion.

The model has received support from an experiment by Humphreys, Allen, and Estes (1968). Lists of problems were constructed so that rewards anticipated by *S* were different from the rewards they actually received during training. The choices of *S*s proved to be related to the anticipated reward values rather than the rewards actually received.

However, the model has received some criticism. Allen and Estes (1969) examined the assumption that exact values of both rewards are learned by *S* in order to solve the problem. They modified the situation used by Humphreys et al. (1968) by having *S* not only indicate his choice on each trial, but also indicate the reward value he expected. Except for having to make value judgments, the experiment was the same as the one reported in Humphreys et al. The results showed that if *S* learned the correct reward value, he typically did not learn it until after he had begun his terminal sequence of correct choices. Hence, the assumption that reward values are explicitly learned before *S* reaches the final state of the model was not supported.

---

[1] This paper is based on a dissertation submitted to the Graduate School of Stanford University in partial fulfillment of the requirements for the PhD degree in psychology. The author is indebted to W. K. Estes, chairman of the dissertation committee, for his invaluable advice and encouragement. Thanks also to R. C. Atkinson, G. H. Bower, and L. M. Horowitz for their instruction and assistance. The research was supported in part by Grant BG 3878 from the National Science Foundation to Stanford University and United States Public Health Service Grant MH 16100 from the National Institute of Mental Health to Rockefeller University.

[2] Requests for reprints should be sent to Gordon A. Allen, Department of Psychology, Miami University, Oxford, Ohio 45056.

Allen and Estes (1969) presented two possibilities to replace the previous assumption. First, during training, *S* might acquire enough information concerning each reward value to indicate whether the value is low, medium, or high on the scale of available values, but not enough information to specify it exactly. Second, *S* might first acquire a relational bit of information regarding each item so he can specify which alternative is associated with the higher reward value.

EXPERIMENT I

The first experiment was a modification of the one reported by Allen and Estes (1969). The modification was necessary because their design had two limitations. First, on a single trial, *S* indicated only the reward value of the alternative he chose; he was not required to indicate the value of the alternative. Therefore, the data could not reveal instances when *S* learned the reward value for an alternative but did not again choose it. Second, although there was a similarity in the pattern of statistics of the choice data between the Humphreys et al. (1968) and the Allen and Estes (1969) experiments, the data suggested that the additional task of giving value judgments biased *S*s to attend more closely to individual reward values and less to structural aspects of the situation.

The training portion of the experiment was exactly like the situation used by Humphreys et al. (1968) except that recall test trials on the individual stimulus words were inserted at various points during the presentation of the second list. After the presentation of test trials, normal choice trials were resumed. At the conclusion of the series of choice trials, a second block of test trials was presented.

It was not economically feasible to have independent groups of *S*s interrupted after each successive presentation cycle for the presentation of the first block of test trials. Rather, three independent groups of *S*s were used: one group was interrupted after 4 cycles, which is, on the average, before the trial of last error for the choice problems; one group was interrupted after 8 cycles, which is the approximate trial of last error; and one group was not interrupted until after 16 cycles, which is the normal sequence. Although this method did not allow for a trial-by-trial examination of the acquisition of the reward values, it did permit examination of the use of the reward values while controlling for the possible disrupting effects of emphasizing the reward values.

No further work had been done using the FIN procedure since that of Keller et al. (1965). It was unknown whether *S*s learned reward values when they were given FIN. Since this was an experiment on the use of information, it was decided to include three additional groups of *S*s who would receive FIN regarding the reward values, but who would also have test trials inserted after 4, 8, or 16 cycles.

Method

Subjects.—Forty-eight paid *S*s participated in the experiment. All were of college age, and most were associated with Stanford University. Eight *S*s were randomly assigned to each of six conditions by means of a table of random permutations.

Apparatus.—Each *S* sat by himself at a table in a sound-attenuated, air-conditioned room which contained only the equipment necessary to display stimulus materials and to enable *S* to make a response. On a table in front of *S* was a metal box containing two rows of response buttons. At the beginning of the experiment, the row containing two buttons was exposed; the left button was marked "L" and the right one "R." The metal strip which covered the second row of buttons was removed at the appropriate time. In this row there were nine buttons, labeled with the digits 0 through 8.

On a second table in front of *S* were two stimulus display boxes, one on top of the other. Eight characters could be displayed in each box. On choice trials, the trigram pairs appeared in the upper box and the payoffs appeared in the lower display, each directly beneath the vowel of the trigram with which it was paired. On test trials, the trigram stimulus appeared in the upper box and nothing was displayed in the lower box.

TABLE 1

PROPORTION OF SUBJECTS GIVING THE CORRECT REWARD VALUE

Block	Time first block presented					
	4		8		16	
	Relative value		Relative value		Relative value	
	High	Low	High	Low	High	Low
PIN procedure						
First	.44	.38	.56	.31	.60	.48
Second	.68	.49	.75	.45	.61	.46
FIN procedure						
First	.25	.18	.36	.26	.40	.24
Second	.49	.44	.54	.36	.36	.25

Note.—Each proportion was obtained by collapsing over *S*s and reward values; therefore, each proportion is based on 80 data points.

A second room, adjoining the *S* room and sharing a sound-attenuating window with it, contained the automated verbal learning equipment. The equipment was operated by means of a modified IBM 526 summary punch which read the prepunched IBM cards and transmitted the information to the programming unit. The programming unit displayed the materials at the appropriate times and recorded *S*'s responses.

Stimuli and reward combinations.—The stimuli were trigram pairs formed from a pool of 40 CVCs, taken from the set of trigrams which had an association value of 53 in the Glaze listings (Underwood & Schulz, 1960). Since a "D" looked like an "O" on the display, no trigrams with a "D" were used. The set of payoff pairs were 8–0, 8–5, 7–5, 6–5, 5–3, 3–2, 3–1, 3–0, 6–7, and 1–2.

The stimulus-payoff lists were constructed so that one *S* in each condition had the same random pairing of trigrams and assignment to payoff values. For each *S*, the left–right sequence and the presentation order within a cycle were randomized. The presentation order of the test trials was randomized for each *S* and for each block of test trials. The randomization procedure used to generate the stimulus-payoff lists and the presentation order was the same as that used by Humphreys et al. (1968).

Procedure.—At the beginning of the experiment, *S* was instructed that when a trigram pair appeared on the top screen, he was to press the button which was on the same side of the screen as the trigram he believed to be worth more points. It was made clear to *S* that the amount of money he earned was a function of the points he earned.

The trigram pair appeared and remained on until after *S* made his response. When *S* pressed a response key, it was illuminated; after 1.5 sec., the payoff values appeared on the lower screen. The trigram pair, the illuminated key, and the reward value were turned off after 1 sec. The next trial began after an ITI of 1 sec.

Depending upon the condition *S* had been assigned, he was interrupted during the second list. The row of nine buttons was exposed and left uncovered. The instructions indicated that when a single trigram appeared *S* was to press the button which corresponded to the payoff value of the trigram. If *S* had been interrupted after 4 or 8 cycles of choice trials, the first block of test trials was presented and the choice trials were resumed. After a short break, the second block of test trials was given. In the 16-cycle conditions the two blocks of test trials were presented last.

After the last block of test trials, *S* was told the purpose of the experiment, his total number of points, and the amount of money earned. The amount of money earned per *S* ranged from $1.75 to $2.25, with an average of $2.00.

Results

Since the data from the choice trials are not of primary interest in this study, choice trial data are not presented; they are available in Allen (1971). The proportion of *S*s giving the correct reward value of the high members and the low members of all reward pairs in a condition are shown in Table 1.

A four-factor analysis of variance was computed where the dependent variable was the number of *S*s who gave the correct reward value. The four factors were information procedure used, time of initial test blocks, relative value of the reward in a pair, and first versus second test block. The last two factors were repeated within *S*s.

There was a reliable difference due to information procedure, with performance being better using PIN, $F\ (1, 54) = 26.59$, $p < .01$. There was no effect due to the time of initial test block. There was a reliable difference between high and low members of a pair, with more *S*s giving the correct value of the high member, $F\ (1, 54) = 17.38$, $p < .01$. There was also a reliable difference between the two blocks of test trials, with more correct responses on the second block, $F\ (1, 54) = 39.61$, $p < .01$. Only one interaction was significant: the Time × Block interaction, $F\ (2, 54) = 8.77$, $p < .01$. The proportion correct increased over the first block, but remained constant over the second block.

The pattern of results obtained, if one examines the mean estimates of the reward values given by *S*s, was exactly the same as the pattern exhibited by the proportion of *S*s giving the correct reward value. The mean estimates of *S*s who received PIN was closer to the actual value than those who received FIN. When mean estimates for a particular reward value were pooled, regardless of the alternative reward value, there was a monotone relationship between actual values and mean estimates for PIN conditions, but not for FIN conditions.

The results reported above have been for judgments of the reward values of individual stimuli. Of more interest was the distribution of judgments for each reward pair; the proportion of *S*s who made correct judgments for both reward values in a pair, for neither reward value, and for exactly one reward value are shown in Table 2. These proportions are based on the second block, and only those protocols having correct choices on the last two training trials are included. Since the information necessary to solve the choice problem was of interest, a criterion was set. Almost all of the protocols excluded represent *S*s who absorbed on the lower alternative early in training.

From Table 2 it is evident that inserting test trials into the sequence of choice trials increased the learning of the reward values in both information procedures. Also, in neither condition is it typical for *S* to learn both values.

It may be that both reward values were known by *S* when the terminal run of correct choices was begun, but during the terminal sequence of correct choices the reward value of the incorrect choice was forgotten. To test this hypothesis, protocols having only correct choices between the first and second block were analyzed separately. There was no evidence for forgetting; rather, there was a shift from either knowing neither to knowing one, or from knowing one to knowing both.

TABLE 2

PROPORTION OF SUBJECT PAIRS MAKING ZERO, ONE, OR TWO CORRECT JUDGMENTS DURING THE SECOND BLOCK OF TEST TRIALS, GIVEN THAT THEIR TWO LAST CHOICES WERE CORRECT

Cycle	*N*[a]	No. of correct judgments		
		0	1	2
PIN procedure				
4	68	.16	.48	.35
8	73	.16	.44	.40
16	70	.21	.51	.27
FIN procedure				
4	77	.40	.29	.31
8	77	.38	.31	.31
16	77	.56	.26	.18

[a] Of the 80 possible, obtained by collapsing over reward values and *S*s, *N* is the number meeting the criterion that the last 2 choices were correct.

The reward values for a problem provided not only information concerning the exact value of each reward, but also the alternative which had associated with it the higher value, independently of the absolute value of either reward. The *S*s might have been learning this relational bit of information. A measure of relational learning is the proportion of protocols exhibiting the correct ordering of a pair of reward values. A correct ordering occurred when *S*'s estimate of the high member was greater than the low member estimate, regardless of the correctness of either estimate. The protocols without correct orderings were partitioned into two groups, those with incorrect orderings and those with equal orderings. These three statistics are presented in Table 3.

The pattern of results was quite similar for both information procedures. The proportion giving correct orderings on the first block increased as the time of initial test increased, and the proportion correct on the second block was approximately equal for all conditions.

Differences become apparent when choice trials are considered along with orderings. An incorrect absorption is a protocol in which at least the last five choices are errors. The proportion of incorrect absorptions for the 4-, 8-, and 16-cycle PIN

TABLE 3

Proportion of Subject Pairs Which Have Correct, Incorrect, or Equal Orderings of Reward Values

Time first block presented	Block					
	First ordering			Second ordering		
	Correct	Incorrect	Equal	Correct	Incorrect	Equal
PIN procedure						
4	.64	.26	.10	.87	.10	.09
8	.77	.14	.09	.82	.09	.09
16	.81	.11	.08	.84	.12	.04
FIN procedure						
4	.66	.15	.19	.85	.05	.10
8	.75	.16	.09	.86	.10	.04
16	.79	.11	.10	.76	.12	.11

Note.—Each proportion is based upon 80 data points.

condition are .09, .04, and .08, respectively. It would be expected that *S*'s ordering of reward values on the second test block would be incorrect if he had incorrectly absorbed. In fact, there was good agreement between the proportion of incorrect absorptions and the proportion of incorrect orderings for the PIN condition. In FIN conditions, however, there were no incorrect absorptions, but the proportion of incorrect orderings was approximately the same as the PIN condition.

Another difference between the two procedures is the use of relational information by *S*s. If there were perfect utilization of relational information, the probability of making a correct choice (CC) on the first trial following a block of test trials should be equal to 1, if *S* had the correct ordering (CO) for his estimates, and it should be equal to 0 if *S* had the incorrect ordering (IO). In the 4-cycle PIN condition this prediction was almost realized: $P[CC|CO] = .88$ and $P[CC|IO] = .24$. In the 8-cycle PIN condition, both conditional proportions increased: $P[CC|CO] = .94$ and $P[CC|IO] = .45$. There was no relationship manifest for all of the FIN conditions; all of the conditional proportions were close to 1.

Experiment II

From the results of the previous experiment and Allen and Estes (1969), it is clear that it is not necessary for *S* to learn the exact value of either alternative in order to solve choice problems under either PIN or FIN procedures. The function of the reward values is not clear; however, three possibilities immediately present themselves. First (Hypothesis 1), reward values provide information as to the correct response to an item, what is learned being the direction of an inequality. In this case, the size of the difference between the rewards, ΔR, would not necessarily affect the rate of acquisition of this bit of information. Second (Hypothesis 2), reward values influence the rate of acquisition of differential approach–avoidance tendencies to the component stimuli of an item; in this case, it would be expected that the rate of acquisition would vary directly with ΔR. The third possibility (Hypothesis 3) is that partial information regarding each reward value is acquired by *S*; the value of the reward is not remembered, but rather the value is encoded by *S* in some manner, and that code is remembered.

The first two explanations, Hypotheses 1 and 2, are similar in that each problem has an alternative which is a "winner," i.e., the correct alternative, and a "loser," the incorrect alternative. Whether a particular reward value is a winner or a loser does not depend upon the absolute value of the reward; rather, it depends entirely upon the relative values of two alternatives. Thus a reward of 7 may be a winner in one problem and a loser in another problem.

The two hypotheses differ in the assumed process by which rewards indicate which alternative is the winner. Hypothesis 1 assumes that learning the inequality, i.e., which alternative is the winner, is an all-or-none process, independent of ΔR. Hypothesis 2 assumes that an approach response is learned to both stimuli in a pair, but the rate of acquisition is faster the larger the reward associated with a stimulus. The rate of acquisition of differential choice is thus a function of ΔR.

The third explanation, Hypothesis 3, is quite different from the first two in that information concerning the value of each reward is learned, and this information is related to the absolute value of the reward. A particular reward value, such as 7, would be encoded in a particular way, regardless of whether the response associated with it is the correct response. This explanation is a generalization of the corresponding assumption in the model presented by Keller et al. (1965). It still requires the assumption that at the moment of choice, *S* scans the possible alternatives and associated outcomes. It also maintains the all-or-none assumption governing the learning of associations.

Experiment II was designed to distinguish among the three hypotheses. The experimental technique combined the training procedures of Exp. I with transfer testing by the method of paired comparisons (PC). After *S* was trained on a set of problems, he was presented with a new set of problems which were generated by making new pairs of the component stimuli of the training pairs. The *S*'s task was again to choose the alternative which would yield more points. The pairs of most interest in the PC task were those in which a reward that was a winner during training was paired with a reward that was a loser, but with the loser having a higher absolute value than the winner.

The method of PCs would generate too many pairs to complete during an hour if used without modification; therefore, the set of reward values used in the training phase of the experiment was constructed so that a reasonable number of pairs in the PC task would be sufficient to differentiate the models. The list was constructed so that the probability of a winner varied directly as the value of the reward except for two reward values. The value 5 was always a loser and the value 4 was always a winner. Another feature of the list was that all pairs of reward values with a ΔR of 1 are included except for the pair 5–4, and the pairs which involved the extremes of the reward value range, 9 and 0. The values 9 and 0 were paired together and defined the range of reward values; all other reward values appeared at least three times.

Using this set of training pairs, it is possible to form several useful sets of PCs. First, there is the set of all possible combinations of rewards which were in training pairs that had a ΔR of 1. For example, using the training pairs 8–7 and 7–6, the PC pairs 8_7–7_8, 8_7–6_7, 7_8–7_6, 7_8–6_7, and 7_6–6_7 were constructed. The subscript of the reward value indicates the reward with which the item was paired during training.

The second set of PCs involve the values 5 and 4. There are three subsets: the set comparing 5 with 4, e.g., 5_6–4_3; the set comparing 5 with a winner of value less than 5, e.g., 5_6–3_1; and the set comparing 4 with a loser of value greater than 4, e.g., 4_3–6_7. The third set of PCs involve equal reward values, e.g., 3_2–3_4 and 6_7–6_5. The fourth set of PCs are the set of training pairs.

The PC trials provide evidence regarding the learning of relationships among the component stimuli. The technique does not allow for the direct assessment of *S*'s knowledge of the reward values. Therefore, after the sequence of PC trials, test trials were presented during which *S* was required to give an estimate of the reward value associated with each stimulus. Perhaps *S*'s judgments would differ depending upon whether both members of a pair were presented or only one member, so both methods of presentation were used.

All of the experimental conditions included the three types of test trials mentioned above. During training, the two types of information feedback were given to independent groups of *S*s. The amount of training was also manipulated; groups of *S*s received either 4, 8, or 16 cycles of the training list.

Method

Subjects.—Sixty-three paid *S*s participated in the experiment. The *S*s were from the Rockefeller University community, nearby high schools and colleges, and were readers of local newspapers which carried classified advertisements. Because of a programming error, the data of the first 3 *S*s were discarded.

Apparatus.—Each *S* sat by himself in an air-conditioned room containing a Techtronix oscilloscope and two response panels. At the beginning of the experimental session, only the 2-button response box was visible; the buttons were unmarked. At the appropriate time, the 10-button response panel was put on the desk; the 10 buttons were marked with digits 0 through 9. A black plywood board covered the front of the oscilloscope except for the screen. Each character, formed by dots of light, was 1 cm. high (six dots) and up to .5 cm. wide (five dots). Rewards were displayed beneath the middle character of the syllable with which they had been paired.

A PDP-8/I computer was in the adjoining room. Once *E* had set the switches to indicate the experimental condition which was to be used, the experiment was under computer control. The sequence of events during each trial was determined by the computer program, and the data were recorded by the computer.

Stimulus and reward combinations.—The stimuli for the practice trials were word pairs formed from a pool of 10 common, three-letter English nouns. The reward pairs used were 9–0, 8–1, 9–8, 8–0, and 1–0.

The stimuli for the training trials were trigram pairs formed from a pool of 32 CVCs which had an association value of 47 or 53 in the Glaze listings (Underwood & Schulz, 1960). The set of trigrams was so constructed that all vowels occurred with approximately equal frequencies and no sequence of two letters was duplicated in the list of trigrams. The reward pairs used in the training list were 9–0, 7–3, 6–2, 8–5, 6–3, 4–1, 8–6, 7–5, 4–2, 3–1, 8–7, 7–6, 6–5, 4–3, 3–2, and 2–1.

The same randomization techniques used in Exp. I were used to form the stimulus-reward combinations and the trial sequence; the only difference was that complete randomization was done for each *S*.

The stimuli for PC trials were pairs of trigrams which were associated with the list of reward pairs used in the PC block. The side on which the lower valued trigram appeared was randomized, as was the order of presentation.

The stimuli for the first block of judgment test trials consisted of each of the trigram pairs used in the training block, and an arrow which was located beneath one of the trigrams and pointed to it. Each pair appeared twice within the block of trials, but on each trial the arrow pointed toward a different trigram. The side on which the arrow appeared was randomized, as was the order of presentation within the block. The stimuli for the second block were the trigrams presented singly. Each trigram appeared once during the block, and the order of presentation was randomized.

Procedure.—The *S*s were randomly assigned to the experimental condition by means of a table of random permutations. The instructions read to *S*s were equivalent to those used in Exp. I.

The sequence of events for both the practice and training choice trials was as follows: the stimulus appeared and remained on until *S* made his response. When *S* responded, the appropriate reward values appeared on the screen beneath the stimuli and remained on for 2 sec. The ITI was 1.5 sec. If no response was made before 15 sec., the word "ZERO" appeared on the screen for 2 sec.

After the block of practice trials, *S* was told the number of errors he made during the last cycle of the practice list. Any questions were answered by paraphrasing the instructions.

After the block of training trials was presented, *S* was instructed that trigrams used in training would be shuffled and new pairs formed. The procedure was the same as training trials except that after *S* made his response, the stimulus remained on the screen for 1 sec. No new information was given to *S*. The maximum time *S* had in which to respond was 15 sec.

After the PC trials, the 10-button response panel was placed in front of *S*, and the instructions indicated that when the stimulus appeared on the screen, *S* was to recall the number of points associated with the trigram which had the arrow beneath it. The *S* had 15 sec. to respond. After *S*'s response, the trigram pair and arrow remained on the screen for 1 sec. The ITI was 1.5 sec. The block of judgment test trials on individual trigrams was presented after the completion of the first block of judgment test trials. The sequence of events for the trials was the same as the first block except single trigrams were the stimuli.

There were about five instances during the PC and judgment blocks in which *S* failed to respond within the allotted 15 sec. In this event, *E* wrote the stimuli on a sheet of paper, and after the block of trials was completed he had *S* make an appropriate response on the paper.

After the experiment was over, *S* was told the purpose of the experiment and his number of points, which was tallied by the computer, and was given the amount of money he had earned. The money earned per *S* ranged from \$1.75 to \$3.75, with an average of \$2.25.

Results

Again, data from the choice trials are not presented. The first set of PCs were the pairs used in the training list. The overall proportions of correct responses for 4-, 8-, and 16-cycle, PIN conditions were .60, .66, .81, respectively; and the proportions correct for the 4-, 8-, and 16-cycle FIN conditions were .74, .90, and .84, respectively. While the proportion correct clearly increased as training increased for the PIN condition, the 16-cycle FIN group performed slightly worse than the 8-cycle FIN group.

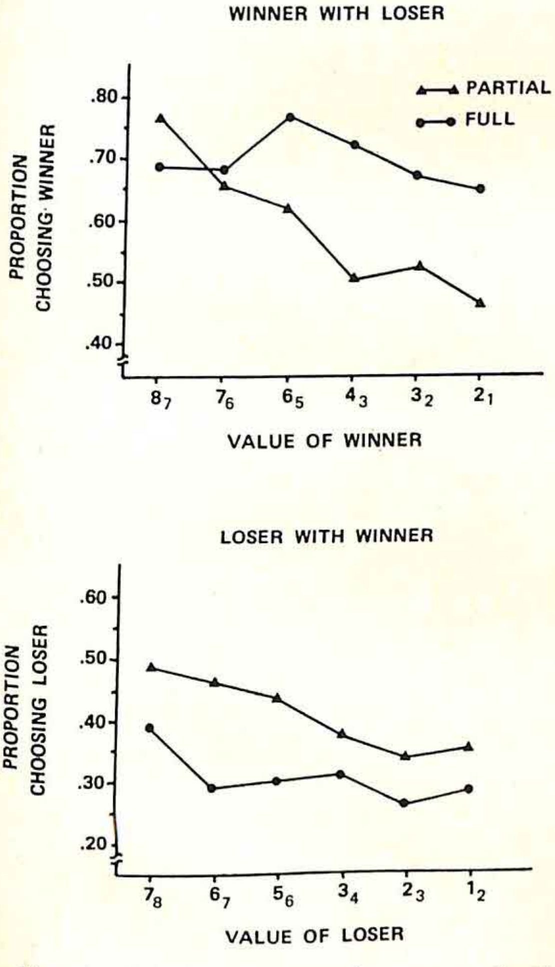


FIG. 1. Paired comparisons between a winner and a loser both of which came from training pairs with $\Delta R = 1$.

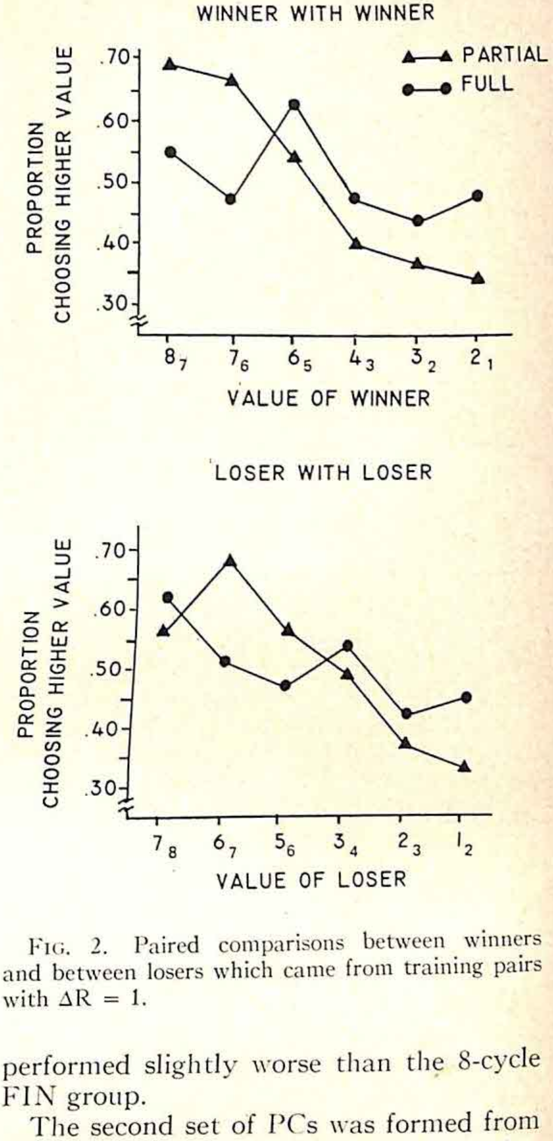


FIG. 2. Paired comparisons between winners and between losers which came from training pairs with $\Delta R = 1$.

The second set of PCs was formed from rewards coming from pairs which had a ΔR of 1 on training. For presentation here, these data were partitioned into comparisons between winners and losers, and pooled within PIN and FIN conditions. First, the proportion of times that a particular winner was chosen over the various losers was calculated. For example, the winner 6_5, in the PIN condition, was chosen 61% of the time over the losers 7_8, 6_7, 5_6, 3_4, 2_3, and 1_2. The proportions are presented in the top graph of Fig. 1. The 95% confidence interval is .43 to .57, assuming that the choices are equally probable and independent. The proportion decreases monotonically as the value of

TABLE 4

Proportion of Times a Loser Was Chosen over a Winner When the Winner Was of Lower Value

Training trial	PC pair		
	High loser with 4	5 with 4	5 with low winner
PIN			
4	.68	.57	.64
8	.50	.48	.51
16	.38	.34	.31
FIN			
4	.39	.49	.34
8	.23	.34	.21
16	.29	.30	.30

Note.—Each entry is based upon 90 data points.

the winner decreases for the PIN conditions. The FIN curve is essentially flat and well above chance.

The proportion of times each loser was chosen when it was paired with a winner is shown in the bottom graph in Fig. 1; for example, the loser 3_4, in the FIN condition, was chosen 31% of the time when it was paired with one of the winners, 8_7, 7_6, 6_5, 4_3, 3_2, and 2_1. The proportions for the PIN conditions decrease monotonically as the value of the loser decreases, indicating that small reward values were avoided more than large values. The entire curve for the FIN conditions is below chance, indicating a marked avoidance of losers, and there is not a strong relationship with respect to reward value.

The proportion of times a particular winner was chosen when paired with another winner is plotted in the top graph of Fig. 2 for the pooled PIN and pooled FIN conditions. All reward values came from training pairs which had a ΔR of 1. The 95% confidence interval, assuming chance responding, is .42 to .58. The proportions decrease monotonically as the reward value decreases for the PIN conditions, while the proportions fluctuate around the chance level for the FIN conditions.

The proportion of times a particular loser was chosen when paired with another loser is plotted in the bottom graph of Fig. 2. Again data were pooled within information procedures, and all reward values came from training pairs which had a ΔR of 1. The proportions for the PIN conditions decrease as the value of the loser decreases; there is one aberrant point which makes the curve nonmonotonic. The FIN curve decreases only slightly as the reward value decreases.

Three sets of PCs involve reward values 5 and 4; all of them pair a loser with a winner of lower value, e.g., 5_6–4_3. There is no evidence of a relationship between performance and the ΔR of the training pair from which the rewards came; therefore, the results are reported in terms of the proportion of times the loser was chosen per set of comparisons for each experimental condition, Table 4. If chance responding is assumed, the 95% confidence interval is .40 to .60.

In the PIN conditions, the proportions of times a loser was chosen, even though it was of higher value than the alternative, decreased as the number of training trials increased. In the FIN conditions, the proportions are below the chance level.

There are two sets of PCs which involve equal reward values, comparisons among 3's and comparisons among 6's. Within neither set of comparisons is there any evidence of a relationship between per-

TABLE 5

Proportion of Choices for a Loser in Paired Comparisons Involving Stimuli of Equal Reward Value

Training trial	Procedure			
	PIN		FIN	
	Pair		Pair	
	Same[a]	Different[b]	Same	Different
4	.61	.57	.44	.28
8	.48	.43	.39	.16
16	.46	.25	.49	.21

[a] The stimuli were either both winners or both losers on training; the one which was paired with the higher valued alternative was arbitrarily defined as the loser. Each entry is based upon 80 data points. The 95% confidence interval is .39 to .61.

[b] The stimulus was composed of a loser and a winner. Each entry is based upon 120 data points. The 95% confidence interval is .41 to .59.

formance and the ΔR of the training pairs from which the rewards were drawn. The data have been dichotomized into comparisons between a loser and a winner, and between either two winners or two losers. The pair 3_6–3_2 is an example of two losers. The two sets have been pooled. In a comparison involving two losers or two winners, the reward which had been paired with the higher valued alternative on training is arbitrarily denoted as a loser; for example, for 3_6–3_7 the stimulus associated with the 3_7 is considered the loser. The results are presented in Table 5.

In the PIN conditions, for different pairs, the proportion of *S*s choosing the loser decreases monotonically from chance as the amount of training increases. This trend is much less marked for the same pairs.

In the FIN conditions, the proportion of *S*s choosing the losers deviates little from the chance level; while, for different pairs, the proportion fluctuates around a point well below chance, indicating a strong preference for a winner, even though both reward values are equal.

The proportion of *S*s who gave the exact reward during the two blocks of test trials did not differ in the PIN conditions, $t = .45$, $p > .10$. For the FIN conditions, the difference was significant, $t = 2.36$, $p < .05$, but the performance difference was so small, .03, that it was considered of no consequence. Therefore, the data for the two blocks were pooled.

The distributions of judgments for each reward pair used during the training trials are presented in Table 6. Shown are the proportions of *S*s who made correct judgments for both reward values in a pair, correct judgments for neither reward, and one correct judgment and one incorrect judgment.

A two-factor analysis of variance was conducted in which the dependent variable was the average number of *S*s who knew neither reward value for each reward combination. The two factors are the type of information given to *S* and the number of training cycles. There is a reliable difference between the PIN and FIN conditions, with fewer *S*s knowing neither reward value in the PIN conditions, $F(1, 90) = 5.77$, $p < .05$. The main effect of amount of training is highly significant, with the number of *S*s knowing neither reward value decreasing as training increases, $F(2, 90) = 10.01$, $p < .01$. Of *S*s who learned the value of at least one reward value, most *S*s learned the value of exactly one reward. However, of *S*s knowing at least one, the proportion of *S*s knowing both values increases as the amount of training increases; this is true for both PIN and FIN.

TABLE 6

PROPORTION OF SUBJECT PAIRS MAKING ZERO, ONE, OR TWO JUDGMENTS

Training trial	No. of correct judgments		
	0	1	2
PIN procedure			
4	.68	.30	.02
8	.65	.30	.04
16	.46	.43	.10
FIN procedure			
4	.68	.27	.04
8	.72	.24	.04
16	.61	.29	.10

Note.—Each proportion is based upon 160 data points.

DISCUSSION

The two methods of presenting reward information during the choice trials exerted different demands upon *S*; and in response, *S* performed differently in the two types of situations. The rate of learning during the choice trials was much faster under the FIN procedure than under the PIN procedure. However, the difference between the two procedures was not only in the rate of learning. Different types of information were acquired by *S* in the two situations. The present two experiments demonstrated that under neither procedure was the acquired information simply the explicit value of the reward. Most *S*s could not recall the explicit value of both rewards, and, when they were able to recall a value, there was good evidence that the value had been learned after the choice problem had been solved. This view that the type of information used by *S* is determined in part by the test situation is consistent with the theoretical schema of Atkinson and Wickens (1971).

In the FIN conditions, there was nothing in the choice task which required *S* necessarily to learn the exact value of the reward assigned to each stimulus alternative. Following each response, both outcomes were presented and *S* could observe which alternative was worth more points. The minimum information which must be extracted from the outcome event is identification of the winner. It appeared that the strategy of *S* was to acquire the minimum information. During a later phase of the experiments in which the winner of one problem was paired with the loser of another problem, *S* consistently chose the winner, even when the loser was of higher value. If two winners were paired together and presented to *S*, there was no preference for the higher valued one; the same was true when two losers were paired together. If two alternatives which had equal value, but only one of which was a winner during the training phase, were paired together and presented, the winner was consistently chosen. There was no evidence that the preference for a winner was related to the difference between the values of the winner and the loser.

Of the three hypotheses concerning the function of rewards, Hypothesis 1 was clearly supported by data obtained using the FIN procedure. This hypothesis predicts that the winner of a problem during training should be chosen when paired with a loser, independently of the values of the winner and loser in the PC problem, and independently of the difference between reward values in the training problem.

The hypothesis was supported by the PC data obtained from Exp. II and was consistent with the data obtained from Exp. I. In the latter experiment, judgment trials, on which *S* indicated the reward value of each stimulus word, were inserted at various times during the choice phase of the experiment. Only a small proportion of *S*s learned the explicit value of the rewards, but a high proportion learned the relation between the rewards. That is, the recalled value of the winner was typically not correct, but it was higher than the recalled value of the loser.

Hypothesis 1 predicts that all *S*s should learn the correct relation. While a large number of *S*s did learn, a substantial number either had the relation reversed or evinced an equality relationship. The easiest explanation of this finding was that *S*s generated a response appropriate to the learned inequality instead of recalling a reward value; that is, a winner would be given a higher reward value and a loser a low one. This explanation was supported by a number of observations (Allen, 1971). First, there was not a clear relationship between judgments and actual reward values, except for the extreme values. Second, there was considerable fluctuation in the judgments between the two cycles of judgment test trials when both cycles were given at the end of the choice sequence. And third, a reward value which was not included in the set of rewards was emitted a substantial number of times during the judgment test blocks.

The PIN procedure was a more complex situation. While the simplest assumption is that *S*s learn the explicit reward values, and then make choices in accordance with them, this notion was not supported by the data. Both Exp. I and II showed that most of the *S*s learned only one of the outcomes, and that this learning occurred largely after the start of the terminal string of 100% correct responses.

There was substantial evidence that *S*s acquired more information concerning the reward values under the PIN procedure than under the FIN procedure. On the judgment test trials inserted at various points in the series of partial information choice trials, there was a monotonic relation between actual values and judged values of the rewards, and the correlation increased with number of previous choice trials. The reward value which was not included in the reward structure occurred on test trials given early in the choice sequence, but very infrequently on later test trials.

On the PC test trials, there was strong evidence that following PIN training, choice was related to the value of the reward and was independent of whether the reward was a winner or loser during training. In pairs which had unequal alternatives, with one being a winner and the other a loser on training trials, the probability of the winner being chosen decreased as the value of the winner decreased; likewise, the probability of the loser being chosen decreased as the value of the loser decreased. The relationship between preference and value was also present in comparisons between two winners and between two losers. In PC pairs which had two equal outcomes, with one a winner and the other a loser during training, there was no preference for either alternative.

Hypothesis 3 was generally supported by the data obtained with the PIN procedure. There was no evidence from the judgment data that the amount of information acquired concerning an outcome was a function of ΔR, the difference between the two reward values of a pair.

The frequency of error (choices of the lower valued alternative) during learning under the PIN procedure was a function of ΔR. However, this result may be interpreted as a performance effect. Following Keller et al. (1965), it may be assumed that after one member of a pair of alternatives has been encoded and the code stored in memory, the probability of shifting from that alternative on a subsequent choice trial is inversely related to the known reward value.

There was one attribute of the PC data which did not support Hypothesis 3. As the amount of training increased, the relation between preference and reward value was attenuated and an increasing preference appeared for the alternative which was the winner during training, whenever a winner was paired with a loser. If a winner was paired with a winner, or loser with loser, the relation between preference and reward value increased as training increased, as predicted by Hypothesis 3. A possible explanation was that after the choice problem had been solved, and *S* consistently chose one alternative, the favorable outcome was reencoded by *S* as a winner. On transfer pairs which involved a winner and loser, the more recent encoding would be used, since no conflict would arise; on transfer pairs involving two winners, the more recent encoding would not provide a basis for choice, and *S* would revert to the earlier encoding.

Taken together, the results of the present study obtained with both the PIN and FIN procedures support the common basic conceptions of several current theories of reinforcement in human learning (Atkinson & Wickens, 1971; Buchwald, 1969; Estes, 1969; Nuttin & Greenwald, 1968). All of these theories share the assumption that *S* acquires information about relations between alternative choices and rewards rather than having response tendencies directly and automatically strengthened by rewards as in classical law of effect doctrine (Postman, 1962; Thorndike, 1935).

At the same time, some of the more specific aspects of the informational theories are seen to require revisions. All share the assumption that *S* learns relations between stimuli and rewards essentially in paired-associate fashion, the nature of the learning process and the connection between learning and choice being spelled out in greatest detail in the model proposed by Keller et al. (1965) and elaborated by Humphreys et al. (1968) and Medin (1969). The present study demonstrates that this assumption requires substantial revision. Information concerning stimulus–reward relationships is indeed acquired by *S* under both of the procedures studied, and the rate of acquisition is probably independent of reward values. However, the result of learning is not necessarily a single association between an object of choice and a reward value; and the character of the information acquired is determined in part by aspects of the training and test procedures.

REFERENCES

Allen, G. A. Memory probes during two-choice, differential reward problems. Unpublished doctoral dissertation, Stanford University, 1971.

Allen, G. A., & Estes, W. K. Acquisition of correct choices and value judgments in binary choice learning with differential rewards. In, *Communications in mathematical psychology*, Technical report series. New York: Rockefeller University, 1969.

Atkinson, R. C., & Wickens, T. D. Human memory and the concept of reinforcement. In R. Glaser (Ed.), *The nature of reinforcement*. New York: Academic Press, 1971.

Buchwald, A. M. Effects of "right" and "wrong" on subsequent behavior: A new interpretation. *Psychological Review*, 1969, **76**, 132–143.

Estes, W. K. Reinforcement in human learning. In J. Tapp (Ed.), *Reinforcement and behavior*. New York: Academic Press, 1969.

Humphreys, M. S., Allen, G. A., & Estes, W. K. Learning of two-choice, differential reward problems with informational constraints on payoff combinations. *Journal of Mathematical Psychology*, 1968, **5**, 260–280.

Keller, L., Cole, M., Burke, C. J., & Estes, W. K. Reward and information values of trial outcomes in paired-associate learning. *Psychological Monographs*, 1965, **79** (12, Whole No. 605).

Medin, D. L. Models for two-choice differential reward learning. In, *Communications in mathematical psychology*, Technical report series. New York: Rockefeller University, 1969.

Nuttin, J., & Greenwald, A. G. *Reward and punishment in human learning*. New York: Academic Press, 1968.

Postman, L. Reward and punishment in human learning. In L. Postman (Ed.), *Psychology in the making*. New York: Knopf, 1962.

Thorndike, E. L. *The psychology of wants, interests, and attitudes*. New York: Appleton-Century-Crofts, 1935.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning*. Philadelphia: Lippincott, 1960.

(Received November 11, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 90–96

SEQUENTIAL MEMORY FOR FAMILIAR AND UNFAMILIAR FORMS[1]

DAVID M. DEL CASTILLO[2] AND WILLIAM E. GUMENIK
University of Toledo

Recall of serial order was tested with familiar and unfamiliar visual forms at presentation rates either too fast to, or slow enough to allow verbal encoding. It was found that when only a single test trial was given following varying amounts of study trials, the slow presentation rate facilitated recall of serial order of familiar forms but not unfamiliar forms for which verbal codes were presumably relatively unavailable. It was further found that when test trials followed each study trial, allowing *S*s adequate opportunity presumably to develop verbal codes for the unfamiliar forms, the slow presentation rate facilitated recall of serial order for both familiar and unfamiliar forms.

Paivio and Csapo (1969, 1971) have shown that increasing the rate of stimulus presentation during study trials adversely affects serial memory for familiar visual forms to a greater extent than for concrete and abstract words. Paivio and Csapo accounted for this in terms of the relative availability of imaginal and verbal memory codes and in terms of the compatibility of the available memory codes with the demands of the sequential memory task.

Paivio and Csapo (1969, 1971) asserted that stimulus information can be stored in either one, or both, of two ways: an imaginal memory code, or a verbal memory code. They further assumed that visual imagery is specialized for parallel processing, in a spatial sense, but not for sequential processing. On the other hand, verbal memory codes are specialized for sequential processing because of the sequential nature of the auditory-speech system. Hence, it follows that verbal codes would generally be more useful in storing sequential information than would imaginal memory codes. A familiar visual form is one that can elicit both types of memory codes. However, verbal coding of a familiar visual form takes longer than for a word. Increasing the speed of presentation of visual forms to a rate so fast that there is insufficient time for the verbal memory codes to be elicited should lead to a decrement in serial memory, since the necessary verbal memory codes for the storage of sequential information would be less available at a high presentation speed. This decrement should be more marked in serial memory of familiar forms than for words because of the greater amount of time necessary to generate the verbal codes in the case of the forms.

The present experiments were a further investigation of the implications of the Paivio and Csapo (1969, 1971) account of the effects of rate of presentation on serial memory for visual form stimuli. In Exp. I, the relative effects of speed of presentation on the serial memory for familiar and unfamiliar visual forms was examined. It was assumed that verbal memory codes are relatively more available for familiar forms than for unfamiliar forms. The fast speed was at a rate that Paivio and Csapo (1969) found to produce reasonably high free recall and recognition with form stimuli, thus insuring discriminability among forms. The slow speed was a rate found by Paivio and Caspo (1969) to further allow implicit verbal labeling of the forms. It was predicted that serial memory for familiar forms would be greater than for unfamiliar forms because of the greater availability of the verbal memory code for familiar forms. However, serial memory for familiar forms should be affected more adversely by an increase in the speed of presentation than serial memory for unfamiliar forms because of the greater availability of the verbal memory code for familiar forms. Actually

---

[1] The authors wish to thank Julian Mingus for his assistance in running *S*s and Kenneth Laytin for his assistance in collating and analyzing the data.

[2] Requests for reprints should be sent to David M. del Castillo, Department of Psychology, University of Toledo, Toledo, Ohio 43606.

serial memory for familiar forms should not differ from serial memory for unfamiliar forms at a presentation rate too fast for the verbal codes to be elicited.

Experiment II investigated whether increasing the amount of experience with the forms, and thereby allowing *S*s an opportunity to produce a verbal code for the unfamiliar forms, would increase serial memory for the unfamiliar forms, thereby diminishing differences between serial memory for the familiar and unfamiliar forms. The further experience with the familiar forms should not affect serial memory for these forms at the fast presentation rate, which presumably does not allow the verbal codes to be elicited.

EXPERIMENT I

Method

Subjects and design.—Two hundred and forty introductory psychology students at the University of Toledo served as *S*s in a 2 × 2 × 6 factorial design, with 10 *S*s per cell. There were 2 levels of stimulus familiarity (familiar and unfamiliar), 2 rates of presentation (fast and slow), and 6 levels of trial presentations (1, 3, 7, 12, 18, and 25). All 24 conditions of the experiment involved independent groups.

Stimulus materials.—Twelve unfamiliar and 12 familiar forms were constructed by cutting silhouettes of the forms out of black construction paper. The unfamiliar items were 6, 12-point, and 6, 16-point random shapes originally constructed by Vanderplas and Garvin (1959) (12-point No., 20, 21, 23, 25, 29, and 30; and 16-point No., 9, 10, 14, 19, 26, 28). The familiar forms were 12 familiar objects (key, sailboat, rifle, airplane, arrow, shoe, car, lamp, toothbrush, fork, chair, and hammer) drawn in silhouette and having similar degrees of shape complexity to the unfamiliar random shapes; i.e., 6 of the forms had 12 points and 6 had 16 points. The familiar forms were first drawn freehand and then reduced to either 12 or 16 points by placing a point at each angle in the drawing. Straight lines were then drawn between successive points and the outline of the form was cut out of black construction paper. This resulted in clearly recognizable shapes which were easily and consistently labeled.

The 24 stimuli were photographed on both 8-mm. black and white movie film, and on 35-mm. high-contrast black and white film. The forms were photographed in such a manner that they occupied approximately the same area on the film format. The 35-mm. negatives were used to make one 3 × 3 in. black and white photographic print of each of the 24 stimuli which, in turn, were mounted on 3 × 3 in. heavy cardboard.

Procedure.—Ten *S*s were randomly assigned in blocks of 24 to each of the 24 experimental conditions. The experiment was conducted in two parts: a series of study trials, followed by a serial recognition test. During the study trials, *S* was seated at a table facing a screen mounted on the wall approximately 4 ft. away. The *S* was instructed that he would see a number of forms which would be presented several times in the same order. He was told to try to remember the order in which the forms were presented. The *S* was then shown either the list of familiar or the list of unfamiliar forms at either the fast or slow rate for either 1, 3, 7, 12, 18 or 25 trials. The fast and slow presentation rates were controlled by the spacing between forms on the 8-mm. movie filmstrip. At both rates, the stimulus was exposed for $\frac{1}{16}$ sec. At the slow rate, each form was followed by seven blank frames, resulting in a $\frac{1}{2}$-sec. interstimulus interval (two items per second). At this rate, it was assumed that the familiar pictures could be named covertly. The fast rate was accomplished by having only two blank frames after each stimulus item and resulted in a $\frac{3}{16}$-sec. interitem interval (5.3 items per second). This rate was assumed to be too rapid for *S*s to name the familiar forms covertly. (See Paivio & Csapo, 1969).

The lists were presented by a Technicolor 200-A, 8-mm. continuous filmstrip projector at a rate of 16 frames per second. Preceding each trial, the word "ready" appeared on the screen for 2 sec. followed by a blank screen for $\frac{1}{2}$ sec. to signal the beginning of a new trial. At the end of each trial, the screen was blank for $\frac{1}{2}$ sec. Thus, the total intertrial interval was 3 sec. During the study trials, the room was lighted only by a 15-w. bulb to allow *E* to view the apparatus.

Following the study trial(s), *S* was given a pack of 12, 3 × 3 in. cards containing the 12 forms he had seen during the study trials and allowed 60 sec. to place them in the correct serial order. The cards were given in a different random order for each *S*.

Results

The mean number of items placed in their correct serial positions as a function of the number of study trials, stimulus familiarity, and speed of presentation are shown in Fig. 1. It can be seen in Fig. 1 that more familiar forms were correctly positioned then unfamiliar forms, that more forms were correctly positioned at the slow presentation speed than at the fast presentation speed, and that the number of forms correctly positioned increased over trials. A 2 × 2 × 6 analysis of variance of these data indicated that the main effects for familiarity, speed of presentation, and study trials, and the Familiarity × Speed and the Familiarity × Study Trials interactions were all statistically reliable beyond

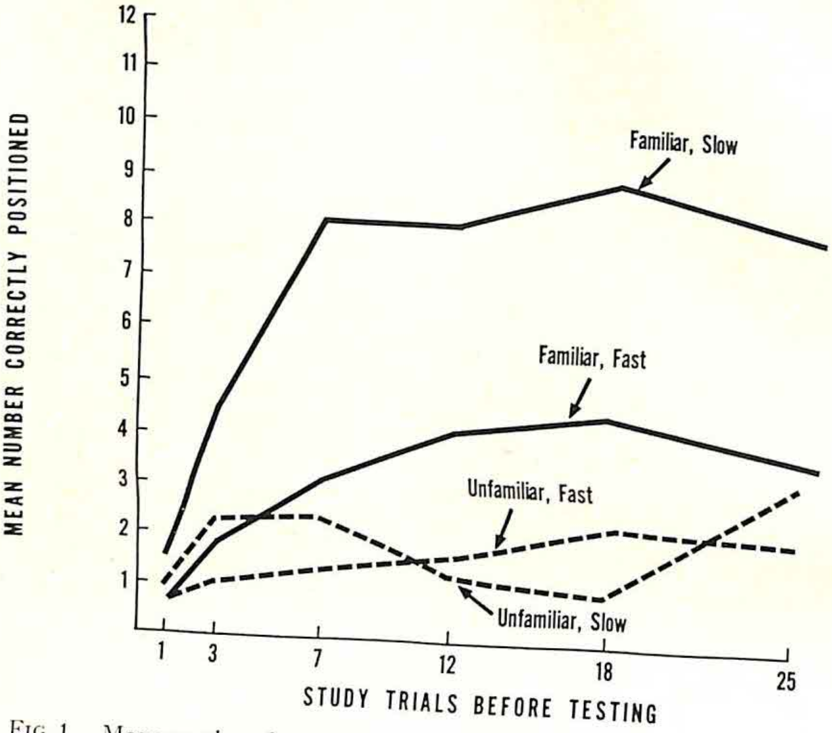


FIG. 1. Mean number of correctly positioned forms in Exp. I as a function of the stimulus familiarity, speed of presentation, and number of study trials before testing.

the .01 level of confidence, F (1, 216) = 185.24, F (1, 216) = 66.22, F (5, 216) = 21.69, F (1, 216) = 64.20, and F (5, 216) = 9.50, respectively. The Familiarity × Speed × Study trials interaction was not statistically significant, F (5, 216) = .25, $p > .05$. Analysis of the simple main effects of familiarity indicated that more familiar forms were correctly positioned then unfamiliar forms at both presentation speeds ($p < .01$). Analysis of the simple main effects for speed at each level of familiarity indicated that speed had a statistically significant effect with the familiar forms ($p < .01$) but not for the unfamiliar forms ($p > .05$).

The significant Familiarity × Study trials interaction reflects the fact that the difference between familiar and unfamiliar forms generally increased with increasing numbers of study trials. Analysis of the simple main effects of familiarity at each of the six trial levels resulted in a reliable difference between familiar and unfamiliar forms at all trial levels ($p < .01$) except the first trial ($p > .05$). The simple main effects for study trials at each level of familiarity revealed a significant study trial effect with both familiar and unfamiliar forms ($p < .01$). In order to determine whether there was any significant increase in performance between the first and last study trials, individual t tests were performed for each of the familiarity conditions at each of the presentation rates using an error term based only on the data being compared. All of the ts were significant beyond the .01 level, except that for the unfamiliar forms at the fast rate ($p > .05$).

The results of Exp. I are in accord with the experimental hypothesis. When the rate of presentation was increased, it was found that performance with the familiar forms was significantly diminished, while that with unfamiliar forms was unaffected. These results are what would be expected if it is assumed that familiar forms are more easily encoded verbally at the slow rate than unfamiliar forms and that increasing the rate of presentation minimizes the amount of verbal encoding that is possible with either type of form.

EXPERIMENT II

As a further test of the hypothesis that the availability of a verbal memory code affects the retention of sequential information, a second experiment was conducted simultaneously with Exp. I. It was reasoned that if *S* was tested for recall of serial order after each study trial rather than after all study trials were completed, as in Exp. I, the test period would provide *S* an opportunity to familiarize himself with the individual test stimuli without providing any new correct sequence information. In the case of *S*s viewing the unfamiliar forms, this would provide an opportunity to develop implicit verbal codes which could be used for the retention of the serial order of the stimuli. This experience should be relatively more facilitative of the unfamiliar forms shown at a slow rate than those shown at a fast rate, since even after *S* had developed a verbal code for the unfamiliar forms it would be unavailable during the study trials given at the fast rate. In contrast, the additional experience with the already familiar forms during testing should have relatively little effect upon the familiar forms since *S*s presumably already had verbal codes for these forms.

Method

The *S*s used in Exp. II were the same as those assigned to the one-trial conditions in Exp. I. Thus, the data for Trial 1 in both experiments were identical. However, following testing after Trial 1, these *S*s were given repeated study trials with the forms, each trial followed by a test for serial recall. The *S*s were given no feedback about the correctness of their ordering of the stimuli on the tests. Study-test trials were continued until *S* had correctly positioned all 12 stimuli on two successive trials or for 25 trials, whichever came first. A perfect score of 12 was given for the remainder of the trials after *S* had reached the criterion. The cards containing the training forms were given to *S* in a different random order on each test trial. In all other respects, the method was identical to that of Exp. I.

Results

The mean number of stimuli correctly positioned in the test for recall of serial order as a function of the number of study-

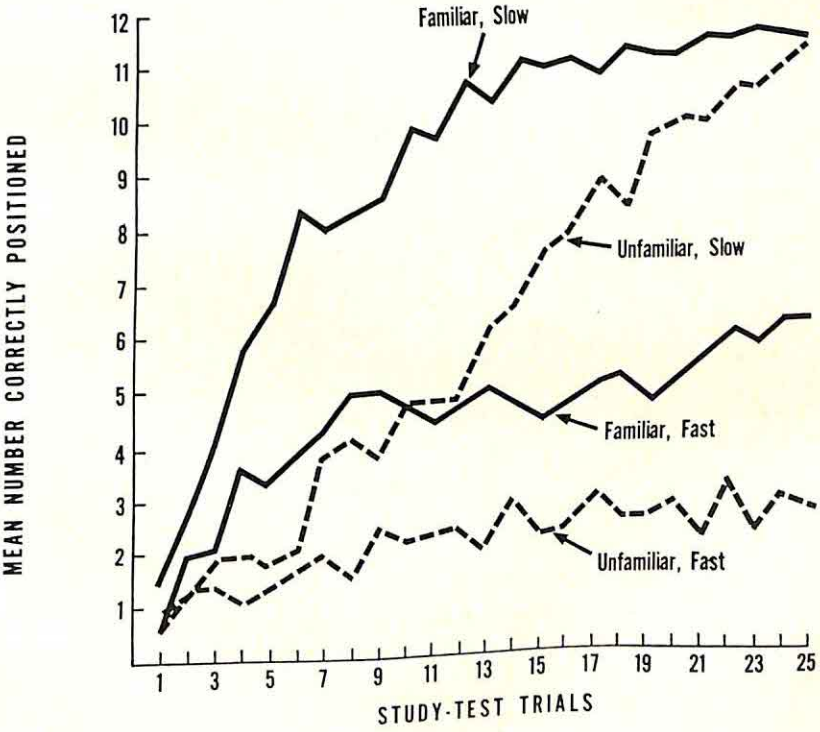


FIG. 2. Mean number of correctly positioned forms in Exp. II as a function of stimulus familiarity, speed of presentation, and number of study-test trials.

test trials, stimulus familiarity, and speed of presentation are shown in Fig. 2. Inspection of Fig. 2 indicates that the pattern of results were very similar to the results of Exp. I, with one major exception: the number of correctly positioned unfamiliar forms presented at the slow rate increases markedly over successive study-test trials. Indeed, performance on Study-Test Trial 25 is nearly identical for familiar and unfamiliar forms at the slow rate (11.3 vs. 11.2 items positioned correctly), although this probably represents a ceiling effect. A 2 × 2 × 25 analysis of variance with repeated measures on the last variable was performed on the results of Exp. II and supports the above conclusions. The analysis revealed statistically significant ($p<.01$) main effects for familiarity, speed, and study-test trials, F (1, 36) = 27.24, F (1, 36) = 72.71, and F (24, 864) = 77.86, respectively. The analysis also revealed significant ($p < .01$) Familiarity × Study-Test Trials, Speed × Study-Test Trials, and Familiarity × Speed × Study-Test Trials interactions, F (24, 864) = 5.07, F (24, 864) = 23.81, and F (24, 864) = 5.51, respectively. The Familiarity × Speed interaction failed to reach statistical significance ($p > .05$), however, this was qualified by its significant interaction with trials reported above. Analysis of the simple effects of trials under each of the combinations of familiarity and speed conditions indicated that trials was a significant ($p < .01$) source of variance under all of the conditions. Individual t tests for correlated means employing error terms based only on the data in each comparison revealed significant ($p < .01$) improvements in performance between the first and last study-test trials for both familiar and unfamiliar forms at both the slow and fast presentation speeds.

In order to make a direct comparison between the results of Exp. I and II, a 2 × 2 × 2 analysis of variance of recall of serial order performance on the twenty-fifth trial in Exp. I and II was performed. This analysis revealed significant ($p < .01$) main effects for familiarity, F (1, 72) = 23.87, speed, F (1, 72) = 86.18, and type of testing procedure (single vs repeated), F (1, 72) = 49.50. The Familiarity × Type of Test Procedure, and Speed × Type of Test Procedure interactions were also statistically significant beyond the .01 level of confidence, F (1, 72) = 60.69, and F (1, 72) = 15.47, respectively. The Familiarity × Speed interaction was not statistically reliable ($p > .05$) although this finding was qualified by a reliable three-way Familiarity × Speed × Type of Test Procedure interaction, F (1, 72) = 10.40, $p < .01$. Analysis of the simple effects revealed that familiarity was a significant source of variance at the slow presentation speed in Exp. I, ($p < .01$) but not in Exp. II ($p > .05$). These results reflect the fact that at slow presentation speeds there was a substantial difference in performance on familiar and unfamiliar forms after 25 presentations followed by a single test for recall of serial order (Exp. I) but no significant difference between them after 25 study-test trials (Exp. II). The analysis of the simple effects of stimulus familiarity also indicated that there was a significant difference in performance between familiar and unfamiliar forms at the fast presentation speed in Exp. II, ($p < .01$), but not in Exp. I ($p < .05$), although the difference in Exp. I was in the same direction. Analysis of the simple effects for type of test procedure indicated that significantly ($p < .05$) more familiar forms under the fast presentation rate and significantly ($p < .01$) more familiar and unfamiliar forms under the slow presentation rate were correctly positioned in Exp. II than in Exp. I ($p < .05$). However, there was no statistically reliable effect for type of study-test procedure with unfamiliar forms at the fast rate ($p > .05$).

DISCUSSION

The results of both Exp. I and II are consistent with the Paivio and Csapo (1969, 1971) hypothesis that sequential memory is dependent upon the availability of a verbal memory code. In Exp. I, the results indicated a strong interaction between stimulus familiarity and speed of presentation which was consistent over increasing numbers of trials. Serial memory for familiar forms was sharply reduced by increasing the speed of presentation, while unfamiliar forms were unaffected by an increase in the presentation rate. Presumably, increasing the speed of presentation reduced the ability of *S*s to verbally encode the familiar forms and hence to store information pertaining to the order of presentation. It is assumed that unfamiliar forms elicited few, if any, verbal codes during the study trials and thus, were unaffected by an increase in the speed of presentation.

An additional finding in Exp. I was that even at the fast presentation rate, performance with familiar forms was superior to that with unfamiliar forms. Several explanations of this finding seem possible. However, one that seems likely is that *S*s given the familiar forms at the fast rate were able to verbally encode a few of the forms by attending to only part of the list. Thus, even though the presentation rate was too rapid to verbally encode the entire list of forms, it may have been possible to encode some of them and thereby retain some of the sequential information.

The results of Exp. II were quite similar to those of Exp. I, with the exception that recall of serial order for the unfamiliar forms presented at the slow rate improved dramatically with increasing numbers of study-test trials. It is assumed that the additional experience with the forms given by testing for recall of serial order after each trial provided *S*s an opportunity to develop and utilize verbal memory codes for retaining serial order information when the presentation was slow enough to allow their occurrence. Performance with the unfamiliar forms presented at the fast rate was quite low and similar to the performance of *S*s in Exp. I with unfamiliar forms presented at the fast rate. Indeed, terminal performance with rapidly presented unfamiliar forms in Exp. II was not significantly different from the terminal performance of *S*s given these forms at the fast rate in Exp. I. Presumably, the fast rate of presentation with unfamiliar forms in Exp. II nullified the effects of any verbal codes developed during the repeated study-test trials.

Terminal performance with familiar forms presented at the fast rate in Exp. II was superior to that with unfamiliar forms at the same speed, despite the hypothesized availability of verbal codes for both types of forms. Possibly, the verbal codes of the familiar forms were elicited more rapidly than those for the unfamiliar forms, because of their greater strength. After all, *S*s given the familiar forms had extraexperimental experience with forms of the same conceptual class in addition to the 25 study-test trials with the experimental forms.

It is possible that the results of the present experiments might be accounted for by some other type of explanation. One possibility is that the difference in performance between familiar and unfamiliar forms in Exp. I is due to the greater discriminability of familiar forms relative to unfamiliar forms. It must certainly be true that the ability to arrange the test forms in the correct serial order depends on being able to discriminate among them when they are presented on study trials. It is also probably true that unfamiliar forms are initially more confusable than familiar forms. If these assumptions are correct, then it would be expected that providing additional experience with the unfamiliar forms by means of repeated study-test trials, as in Exp. II, would have the effect of reducing the confusability of the unfamiliar forms and thereby enhancing recall of serial order. Increasing the speed of presentation would be expected to make the task of discriminating among forms more difficult, especially when the discrimination is relatively weak, which would be the case with the unfamiliar forms.

Studies of stimulus predifferentiation have repeatedly demonstrated the importance of verbal processes in learning to discriminate among unfamiliar visual forms, even when such processes are operating covertly (e.g., Ellis, 1968; Ellis & Homan, 1968; Ranken, 1963). Such studies have shown that verbal labeling facilitates discrimination of form stimuli, and that *S*s will covertly label forms to be discriminated even when not specifically instructed to do so (see Riley, 1962). In any case, whether the verbal processes operate to enhance discriminability or to facilitate serial encoding, or a combination of these effects, the results of the present experiments are consistent with a verbal encoding interpretation.

Another alternative way of accounting for the results of the present experiments might be in terms of the operation of imaginal encoding processes. It might be argued that imaginal codes may be more easily and more rapidly obtained with familiar than with unfamiliar forms. Increasing the speed of presentation would simply increase the difficulty of obtaining imaginal codes, thus making performance with familiar and unfamiliar forms more alike. However, such an explanation begs the question since it has not yet been empirically established that familiar and unfamiliar forms differ in the ease or speed with which imaginal codes are obtained. Nor do there seem to be any a priori arguments for making this assumption. In contrast, in accord with a verbal encoding explanation, the familiar forms in the present experiments were deliberately selected because of the ease and consistency with which verbal labels could be applied to them.

The verbal encoding explanation of the present results may also be supported on the basis of its congruence with other research findings. Numerous investigators have found that visually presented letters and words are most likely to be confused in STM when they are acoustically similar, thus suggesting a verbal encoding of the visually presented stimuli (e.g., Baddedly, 1966; Conrad, 1962; Gumenik, 1969). Of course, these findings may represent a special case of verbal encoding because of strong verbal habits established with these types of stimuli. However, Glanzer and Clark (1963) found a high negative correlation between serial memory for briefly exposed binary (black and white figures) sequences and the length of verbal descriptions given the sequences by their *S*s. They explained this result in terms of the verbal loop hypothesis that *S*s verbally encode visually presented displays and then respond (serial recall) on the basis of the stored verbal code. Colgate and Eriksen (1970) trained *S*s to associate either one- or three-syllable words to a set of nonsense forms, and then presented the set of forms simultaneously in a brief tachistoscopic exposure, followed by a test probe. Their results indicated that more nonsense forms were recalled when the forms had previously been associated with one-syllable words than when they had been associated with three-syllable words. Presumably, this was because of the greater time required to verbally encode the forms when they had been associated with the longer words.

One further explanation of the current results might be that the difference in performance between familiar and unfamiliar forms reflects differences in stimulus meaningfulness, familiar forms being higher in meaningfulness than unfamiliar forms. However, stimulus meaningfulness is usually defined in terms of the extent to which stimuli are capable of eliciting verbal associations (e.g., Ellis, Muller, & Tosti, 1966; Glaze, 1928; Noble, 1952; Pfafflin, 1960) and, indeed, as already mentioned above, was the basis for defining familiar forms in the present experiments. Hence, this interpretation would be consistent with the verbal encoding account given here.

REFERENCES

BADDEDLY, A. D. Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. *Quarterly Journal of Experimental Psychology*, 1966, **18**, 362–365.

COLGATE, R. L., & ERIKSEN, C. W. Implicit speech as an encoding mechanism in visual perception. *American Journal of Psychology*, 1970, **83**, 208–215.

CONRAD, R. An association between memory and errors due to acoustic masking of speech. *Nature*, 1962, **193**, 1314–1315.

ELLIS, H. C. Transfer of stimulus predifferentiation to shape recognition and identification learning: Role of properties of verbal labels. *Journal of Experimental Psychology*, 1968, **78**, 401–409.

ELLIS, H. C., & HOMAN, L. Implicit verbal responses and the transfer of stimulus predifferentiation. *Journal of Experimental Psychology*, 1968, **76**, 486–489.

ELLIS, H. C., MULLER, D. G., & TOSTI, D. T. Stimulus meaning and complexity as factors in the transfer of stimulus predifferentiation. *Journal of Experimental Psychology*, 1966, **71**, 629–633.

GLANZER, M., & CLARK, W. H. Accuracy of perceptual recall: An analysis of organization. *Journal of Verbal Learning and Verbal Behavior*, 1963, **1**, 289–299.

GLAZE, J. A. The association value of nonsense syllables. *Journal of Genetic Psychology*, 1928, **35**, 255–269.

GUMENIK, W. E. Effects of articulatory activity and auditory, visual, and semantic similarity on the short-term memory of visually presented paired associates. *Journal of Experimental Psychology*, 1969, **82**, 70–74.

NOBLE, C. E. An analysis of meaning. *Psychological Review*, 1952, **59**, 421–430.

PAIVIO, A., & CSAPO, K. Concrete image and verbal memory codes. *Journal of Experimental Psychology*, 1969, **80**, 279–285.

PAIVIO, A., & CSAPO, K. Short term sequential for pictures and words. *Psychonomic Science*, 1971, **24**, 50–51.

PFAFFLIN, S. Stimulus meaning in stimulus predifferentiation. *Journal of Experimental Psychology*, 1960, **59**, 269–274.

RANKEN, H. B. Effects of name learning on serial learning and recognition learning with random shapes. *Psychological Reports*, 1963, **13**, 663–678.

RILEY, D. A. Memory for form. In L. Postman (Ed.), *Psychology in the making*. New York: Knopf, 1962.

VANDERPLAS, J. M., & GARVIN, E. A. The association value of random shapes. *Journal of Experimental Psychology*, 1959, **47**, 147–157.

(Received November 19, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 97–101

ATTRIBUTE SELECTION IN CONCEPT IDENTIFICATION [1]

DOUGLAS C. CHATFIELD [2] AND ERWIN J. JANEK

Texas Tech University

One hundred and twenty *S*s were given two concept-identification tasks. To view a particular attribute of an instance, *S* made an overt button-pressing response. It was assumed that the number of attributes selected on each trial was a function of the number of hypotheses in *S*'s focus sample. Results revealed that the number of attributes selected for observation per trial decreased over the criterion-run trials after the trial of last error. The second task was comprised of six trials wherein *S*s received either three errors followed by three successes or three successes followed by three errors (bogus feedback). Results showed a systematic decrease in attribute selection over both positions of the success trials and over the error trials when they were the initial three trials. An increase in attribute selection occurred, however, over the last three trials resulting in errors.

Trabasso and Bower (1968) proposed a model for concept identification in which *S* is viewed as testing several hypotheses simultaneously in search for the solution. On the first trial of a problem and following each trial on which an error is made, *S*s are assumed to randomly select with replacement a focus sample of attributes from the total set of attributes. Locally consistent response assignments are attached to the selected attributes transforming them into hypotheses. The correctness of these hypotheses are then tested on subsequent trials. On trials in which *S*'s responses are correct, hypotheses in the focus that dictated that response are retained, whereas hypotheses which had dictated an alternative response are eliminated from the focus. The complete focus sample is given up on those trials on which an error was made leading to a new sampling of attributes.

A test of the attribute-sampling assumptions was carried out by requiring *S*s to make an overt observing response for each attribute on each trial, similar to that by Rydberg and Arnberg (1969). In using the number of attributes selected and observed per trial as an indication of the size of the focus samples, the following predictions from the Trabasso and Bower model were tested: (*a*) since *S*s condiser several hypotheses concurrently, *S*s should attend to several attributes per trial; (*b*) as it is assumed that incorrect hypotheses are eliminated from the focus sample on success trials, the number of attributes selected on a particular trial should be inversely related to the number of consecutive successes since the last error; (*c*) attribute selection should increase following the trial of last error (since all *S*s are assumed to select a new focus sample at that point) and then proceed to decrease over the success trials contained in the criterion run; (*d*) attribute selection following an error should be independent of any events in previous trials as the model assumes errors to be recurrent events.

METHOD

Subjects.—The *S*s were 120 undergraduate students serving an introductory psychology course requirement of experimental participation.

Apparatus and procedure.—Stimuli consisted of a set of verbal descriptions of imaginary trees. The six binary attributes of the trees and their values were listed on a sheet of transluscent paper and placed on a display board as shown in Fig. 1. The display board was mounted in a small table at which *S* was seated. On each trial, a description of a tree was given to *S* by illumination of the appropriate values for each attribute. As an example, Fig. 1 is a description of a large, smooth-barked tree with yellow blossoms and green fruit. In order to observe the value of a particular attribute, *S* was required to press its corresponding attribute-selection button. In the example in Fig. 1, *S* would have chosen to ob-

---

[1] This research was supported by funds from Texas Tech University. Special thanks are due Debby Brotherton for aid in the collection of data for this study.

[2] Requests for reprints should be sent to Douglas C. Chatfield, Department of Psychology, Texas Tech University, Lubbock, Texas 79409.

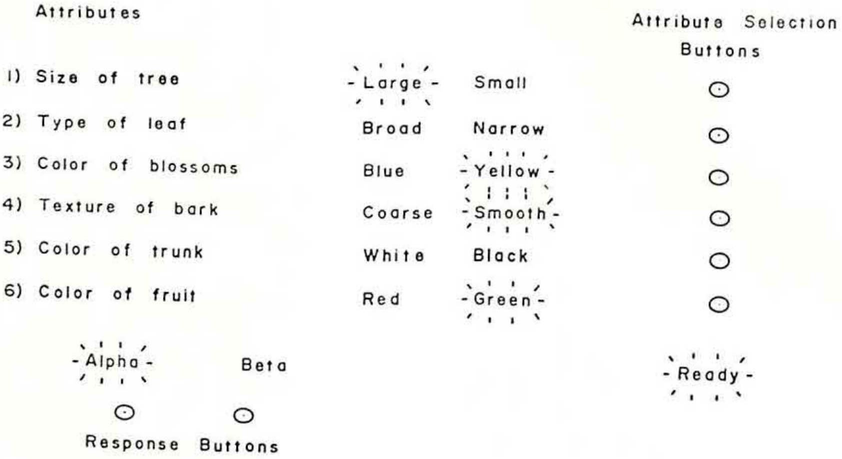


FIG. 1. Schematic representation of the stimulus display.

serve Attributes 1, 3, 4, and 6 and to ignore Attributes 2 and 5. In ignoring Attributes 2 and 5, *S* would not know whether the tree, being described, had broad or narrow leaves, or the color of its trunk. Each *S* was instructed that he could observe as many of the six attributes on each trial as he wished. Following the attribute selection, *S* was required to classify the tree as an alpha or a beta by pressing the appropriate button. This response was immediately followed by 5 sec. of informative feedback via illumination of the correct classification. There was a postfeedback interval of 10 sec. before the next trial as indicated by the onset of a ready light.

Design.—Following the instructions in which the nature of the concept-identification task was described, *S*s were given a practice problem consisting of three binary attributes. For their second task, all *S*s were given a six-attribute problem wherein only one attribute was relevant and the other five irrelevant. The problem was terminated when *S* made eight consecutive correct classifications.

A study by Chatfield and Erickson (1970) had shown that a normal concept-identification problem does not produce enough presolution error and success runs to adequately test the predictions concerning attribute selection over these runs. For this reason, a third task was created wherein error and success runs would be created artificially by using bogus feedback. In this task, *S*s were randomly divided into two groups of 60 *S*s each with the instructions that *E* would randomly resample one of the six attributes to be relevant. For the Error-Success group the feedback given for the first three trials always indicated that *S*s' classifications were in error. The

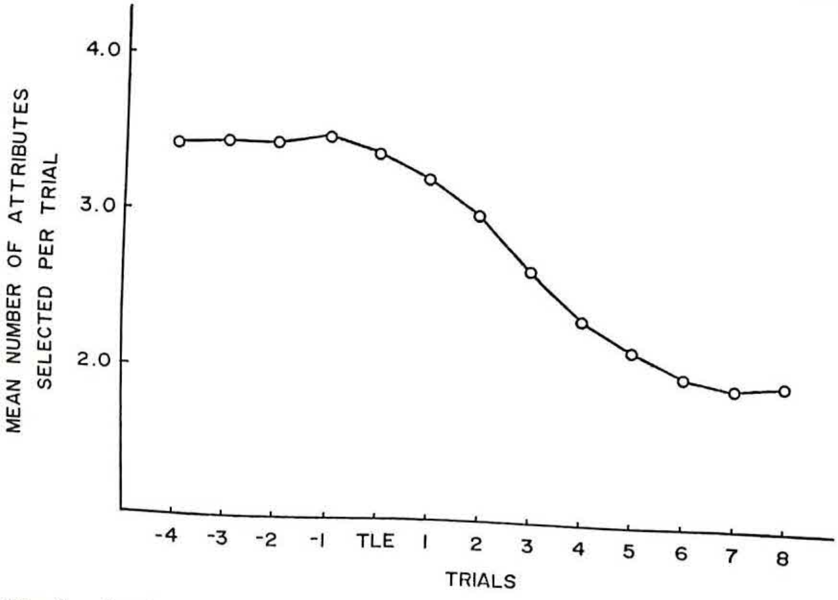


FIG. 2. Attribute selection as plotted backwards from the last trial of the criterion run.

feedback on Trials 4–6 indicated that these classifications were correct. For the Success-Error group, the type of bogus feedback was in the reverse order; i.e., the first three trials always resulted in successes with the next three trials being in error. The third task ended with the six trials thus described.

RESULTS

The major results of *S*'s attribute-selection behavior during the second problem may be found in Fig. 2. In this figure, all of *S*s' protocols were aligned at the trial of last error. It may be noted that *S*s observed an average of three to four attributes on trials prior to their last error and subsequent solution. On the trial immediately following the last error, *S*s decreased their selection of attributes slightly in contrast to prediction (*c* above), that there should be an increase. Attribute selection did continue to decrease over the remainder of the criterion run, however, a finding consonate with prediction (*b* above), since *S* should be eliminating hypotheses over a run of success trials.

The attribute-selection data on the third problem may be logically divided into three sections with respect to the treatment conditions imposed on the two groups of *S*s. There were no treatment differences at the time *S*s selected attributes on the first trial as they had not yet received any feedback. Attribute selection on Trials 2, 3, and 4 should be a function of error (Error-Success group) or success (Success-Error group) feedback on the immediately preceding trials. Finally, attribute selection on Trials 5 and 6 should be influenced by the switch in feedback.

The results for the first six trials of the third problem may be found in Fig. 3. As can be seen, the number of attributes selected in both groups decreased significantly over the first four trials, $F\ (3, 354) = 14.31$, $p < .001$. The Groups × Trials interaction was not significant, $F\ (3, 354) = 1.64$, $p < .10$, but since the decrease in attribute selection over trials for the Error-Success group had theoretical importance, it was analyzed separately and was found to be significant, $F\ (3,\ 177) = 3.54$, $p < .05$. The two groups did not differ significantly in terms of the overall mean number of attributes selected during the first four trials, $F\ (1,\ 118) < 1$.

Trials 4, 5, and 6 were analyzed separately as a 2 × 3 factorial design. Trial 4 was included in this analysis, as it was the change in attribute selection resulting from the new feedback that was of interest. Neither main effect was significant at the .05 level but the Groups × Trials inter-

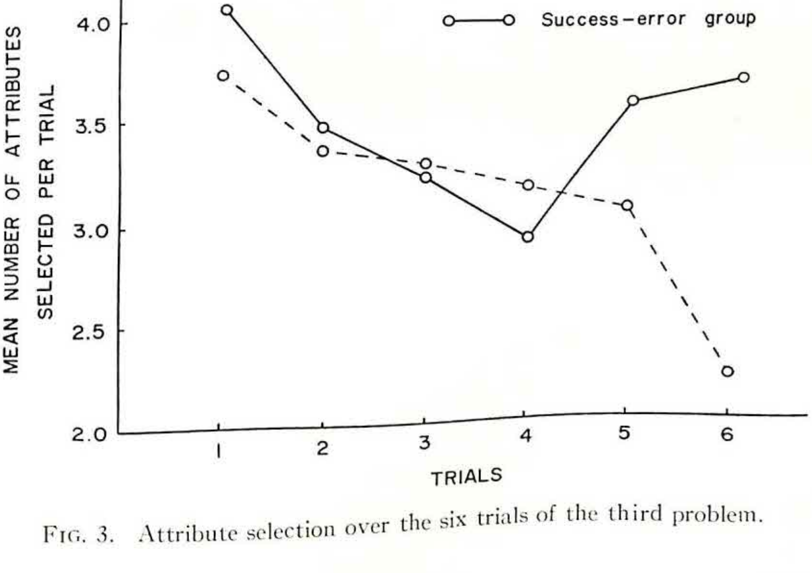


FIG. 3. Attribute selection over the six trials of the third problem.

TABLE 1

Chi-Square Values Reflecting Logical Retention and Deletion of Attributes from Focus As a Result of Feedback Conditions

Item	Error-success group			Success-error group		
	3 and 4	5	6	3 and 4	5	6
Feedback	Error	Mixed	Success	Success	Mixed	Error
Chi-square value	6.89*	.79	7.10*	9.51*	3.36	.03

Note.—Trial number is indicated by the column heads 3 and 4, 5 and 6.
* $p < .01$.

action was; *F* (2, 236) = 14.51, $p < .001$. As can be seen, the Success-Error group increased attribute selection on Trials 5 and 6 while the Error-Success group decreased its selection.

The finding that attribute selection decreased over the intial run of errors but increased over the latter run of errors was unexpected. For this reason it was decided to analyze the data in terms of the specific attributes which were being deleted and retained. The Trabasso and Bower theory would predict that an attribute would not be sampled following a trial on which it was disconfirmed. More specifically, if an attribute was observed to change values over Trials $n - 2$ and $n - 1$ while the feedback remained constant (or vise versa), the attribute should logically be disconfirmed and thus should not be sampled on Trial *n*. For this reason, on each Trial *n*, attributes were classified in two ways: (*a*) whether or not they were disconfirmed in the two preceding Trials $n - 2$ and $n - 1$ (given that they had been observed by *S*), and (*b*) whether or not they were selected on Trial *n*. The resulting frequencies were put into a 2 × 2 contingency table.

As two trials were needed to disconfirm an attribute, only the data from Trials 3–6 could be analyzed. These four trials fall into three feedback conditions. Attribute selection on Trials 3 and 4 would be dependent on feedback conditions on Trials 1 and 2, and Trials 2 and 3, respectively. These first three trials resulted in error feedback for the Error-Success group and success feedback for the Success-Error group. Attribute selection on Trial 5 was dependent on the feedback from Trials 3 and 4, which of course was mixed for both groups. Trial 6 was dependent on the feedback from Trials 4 and 5, which for each group was the opposite of what they had on Trials 1–3. The resulting chi-square values are given in Table 1. As can be seen, *S*s tended to drop disconfirmed attributes and retain those confirmed, as indicated by the magnitude of the chi-square value on trials preceeded by correct feedback. Sampling on Trial 5, preceded by mixed feedback, was relatively unsystematic. Though efficient sampling was found on trials preceded by error feedback early in the problem, it was not found later in the problem as shown by the Success-Error group on Trial 6.

Discussion

Portions of the data upheld the predictions made at the outset and thus supported the model by Trabasso and Bower (1968). There were three places where reduction of the number of attributes selected over a series of successes was shown. The first was during the criterion run of the second problem. The second was evident in Trials 2, 3, and 4 for the Success-Error group in the third problem. The third place was in Trials 5 and 6 for the Error-Success group, also in the third problem. This reduction indicates that *S*s were eliminating hypotheses from the focus sample during those trials. Additional evidence comes from the high chi-square values shown in Table 1, indicating systematic deletion of attributes over success trials. If it is assumed that response latencies are a function of the number of attributes selected, then the present findings are in accordance with the latency data presented by Erickson, Zajkowski, and Ehmann (1966); Erickson and Zajkowski (1967) and Levine (1969).

The attribute-selection behavior following errors was not in agreement with the prediction that *S*s always resample after errors. Attribute selection did increase for the Success-Error group on Trials 5 and 6, indicating that Ss were resampling after their initial series of successes as predicted, but a reduction in attributes selected occurred during the first four trials for the Error-Success group. This would imply that *S*s were eliminating hypotheses during those initial error trials. Furthermore, the significant chi-square value for Trials 3 and 4, given error feedback, show that *S*s were not just randomly dropping attributes from their focus sample. Why attribute selection decreases systematically over an initial series of errors, but increases over errors which had been preceeded by successes, poses a theoretical problem for Trabasso and Bower (1968).

A possible explanation for these findings comes from Levine (1971). In describing his theory of insoluble problems, he proposed that when *S* has exhausted all possible hypotheses, he may either review the set of possible hypotheses again, or begin sampling new hypotheses which are not possible solutions as defined by the experimental instructions. The result of either strategy on the part of *S* would be an increase in attribute selection. This explanation would be applicable to the present study in that it was possible for *S*s to eliminate most all of the possible hypotheses they could have had in their original focus sample by the third or fourth trial. Thus, when told in the latter trials that they were in error, *S*s had no choice but to resample. When the errors came in the first few trials, *S*s had not yet exhausted all the hypotheses in their focus samples and could continue their examination of those remaining.

REFERENCES

Chatfield, D. C., & Erickson, J. R. Dimension selection and memory in concept identification. Paper presented at the meeting of the Midwestern Psychological Association, Cincinnati, April 1970.

Erickson, J. R., & Zajkowski, M. M. Learning several concept identification problems concurrently: A test of the sampling-with-replacement assumption. *Journal of Experimental Psychology*, 1967, **74**, 121–128.

Erickson, J. R., Zajkowsi, M. M., & Ehmann, E. D. All-or-none assumptions in concept identification: Analysis of latency data. *Journal of Experimental Psychology*, 1966, **72**, 690–697.

Levine, M. Latency-choice discrepancy in concept learning. *Journal of Experimental Psychology*, 1959, **82**, 1–3.

Levine, M. Hypothesis theory and nonlearning despite ideal S–R reinforcement contingencies. *Psychological Review*, 1971, **78**, 130–140.

Rydberg, S., & Arnberg, P. W. Selective attending changes during concept learning, extinction, and superstition. *Psychonomic Science*, 1969, **17**, 99–100.

Trabasso, T., & Bower, G. *Attention in learning: Theory and research*. New York: Wiley, 1968.

(Received November 24, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 102–109

AN ONTOGENETIC ANALYSIS OF OPTIONAL INTRADIMENSIONAL AND EXTRADIMENSIONAL SHIFTS[1]

HOWARD H. KENDLER,[2] TRACY S. KENDLER, AND JAMES W. WARD
University of California, Santa Barbara

The tendency to choose an optional intradimensional (ID) and extradimensional (ED) shift was investigated at five developmental levels: preschool, kindergarten, second grade, sixth grade, and college. Over this range, a significantly increasing trend to select an ID shift and a significantly decreasing trend to choose an ED shift were observed. ID *S*s when compared to ED *S*s exhibited higher IQ's and learned both the initial and shift discriminations more rapidly. A comparison between the present optional ID–ED study and a previous optional reversal–ED experiment revealed that the ontogenetic changes are similar. The results are interpreted to be consistent with the coordinated single-unit and mediational S–R formulation.

According to the developmental mediational theory (Kendler, 1971; Kendler & Kendler, 1962, 1968), ontogenetic changes in reversal (R) and extradimensional (ED) shift behavior reflect a transition from a single-unit mechanism in which responses are under direct control of external stimulation to a mediational mechanism in which behavior is controlled by self-generated symbolic cues that represent conceptual categories. Evidence in favor of this formulation comes from developmental studies that show that with age the relative ease of executing an R shift in comparison to an ED shift increases (Tighe & Tighe, 1967) and that the choice of an optional R increases in comparison to an optional ED shift (Kendler & Kendler, 1970; Kendler, Kendler, & Learnard, 1962).

Although Campione (1970) does not question the developmental data just cited, he suggests that other evidence denies the necessity of postulating two modes of behavior. First, Campione cites the inconsistency in experimental results concerning the relative ease with which preschoolers execute an R or an ED shift as being at odds with Kendlers' formulation. Whereas Kendler, Kendler, and Wells (1960) find that preschoolers solve ED more easily than R shifts, other researchers (Dickerson, 1966; Mumbauer & Odom, 1967) report the opposite finding. But evidence about a single-age level, by itself, is irrelevant to the proposition that R shifts, with increasing age, tend to become relatively easier than ED shifts. Because variables other than age influence discrimination-shift behavior (e.g., Kendler, Kendler, & Sanders, 1967), one cannot assume that at a specific age level, such as 4 or 5 yr., an ED shift should *always* be executed more rapidly than an R shift. Developmental changes, not behavior at a fixed age level, are the significant data for the developmental mediational theory.

Second, Campione (1970) suggests that the comparison between intradimensional (ID) and ED shifts is as, if not more, important than the R–ED comparison in evaluating theories of discrimination-shift behavior. Thus he cites as inconsistent with the Kendlers' two-process theory the findings from several independent experiments that ID shifts are easier to execute than ED shifts for rats (Shepp & Eimas, 1964), preschool children (Dickerson, 1966; Mumbauer & Odom, 1967), and adults (Kurtz, 1955). In order to determine whether the presumed superiority of ID over ED shifts for all ages and species extends to an optional-shift design, Cam-

---

[1] This research was supported by National Science Foundation Grant GB 19323. The authors wish to acknowledge the assistance and cooperation of Wayne Pelton, Principal, and Jean Faulk, Sue Hoag, Sue Schaff, and Nancy Winn, teachers of the Kellogg School of Goleta, California; also, Doris Gammons, of the El Montecito Nursery School of Montecito, California. James W. Ward and Doris A. Bregante served as *E*s.

[2] Requests for reprints should be sent to Howard H. Kendler, Department of Psychology, University of California, Santa Barbara, California 93106.

pione compared the behavior of preschool with second-grade children in choosing either an optional ID or ED shift. He found no significant difference; approximately 65% of each group selected an ID shift. The conclusion drawn was that the data are "inconsistent with the notion that there exists an ontogeny of shift behavior [Campione, 1970, p. 299]."

In responding to the second point, the position can be maintained that the ID–ED comparison is not intrinsically superior to the R–ED comparison in reflecting ontogenetic changes. Each comparison, ID–ED and R–ED, emerges from a different theoretical position and thus requires different experimental controls. The ID–ED serves the needs of the mediational attention theory (Zeaman & House, 1963) because it reflects the operation of preshift attending responses presumably uninfluenced by previously learned choice responses. In contrast, the R–ED comparison, in which the discriminanda remain unchanged from preshift to postshift training, possess two advantages over the ID–ED paradigm for the Kendlers' dual-stage theory. First, complex stimulus generalization effects inherent in the ID–ED design and which cannot be simply controlled by counterbalancing, are avoided. Second, the persistence of preshift mediated representational responses to postshift training cannot be impeded by the sudden occurrence of novel discriminanda as is the case in the ID–ED comparison.

In the last analysis, however, an adequate explanation of the ontogeny of discrimination-shift behavior must account for results from both experimental paradigms. In spite of Campione's (1970) results, reservations can still be expressed about his conclusion that developmental changes in optional ID–ED choices do not occur. In addition to the logical problem of proving the null hypothesis is the more serious empirical problem of demonstrating a developmental relationship based on only two points. To evaluate the last criticism, the present experiment was designed to discover whether a sampling of five age groups, ranging from 4½ to 20 yr., would fail to reveal any developmental changes in optional ID and ED choices.

METHOD

Subjects.—A total of 163 *S*s were used divided among five age levels that approximated a geometric progression: mean chronological ages (CA's) of 54.4, 70.7, 94.0, 139.9, and 229.7 mo. These age points were selected to duplicate those used in other developmental studies that employed the optional-shift paradigm (Campione, 1970; Kendler & Kendler, 1970). The lowest age group, the preschool children, were from a private nursery school in an upper-middle-class community, Montecito, California. The next three age levels, kindergartners, second, and sixth graders, were drawn from a public elementary school in a middle-class neighborhood in Goleta, California. The oldest group were college students from an introductory psychology course at University of California, Santa Barbara, who served to meet course requirements. Three *S*s were eliminated from the kindergarten group, 1 because absence prevented completion of the experimental task, and 2 were randomly eliminated from the data analysis because both *E*s inadvertently ran 2 *S*s from the same cell.

Experimental design.—The optional-shift procedure (Kendler & Kendler, 1970), adapted to a comparison between ID and ED shifts (Campione, 1970), was used. Table 1 illustrates the procedure. In the initial discrimination, *S* is confronted with a two-choice discrimination between two pairs of stimulus compounds that vary on two dimensions, form and color, with one (e.g., form) relevant and the other (e.g., color) irrelevant. After reaching criterion, *S* enters the second phase, the shift discrimination, where he learns to discriminate between a new pair of stimulus compounds that vary on the same two dimensions but with new values for each. Both dimensions are now relevant and redundant. After learning to select the correct compound consistently, *S* is shifted immediately to the test series which is designed to discover on what basis the shift discrimination was solved. In the test pair of the third phase, the previously redundant and correct cues, for example, blue and square in Fig. 1, are re-paired and the consistent choice of square indicates an ID choice because the relevant dimension in the initial discrimination was form. The consistent choice of the blue indicates an ED choice. Interspersed between the presentation of the test pair are trials with the relearning pair that continue to be reinforced as in the shift discrimination.

Two sets of stimulus compounds were used that varied in color and form: one with color values of red and green and form values of circle and triangle, and the other with color values of blue and yellow and form values of square and cross. Each *S* learned the initial discrimination with one set and the subsequent shift discrimination and test series with the other. A completely counterbalanced

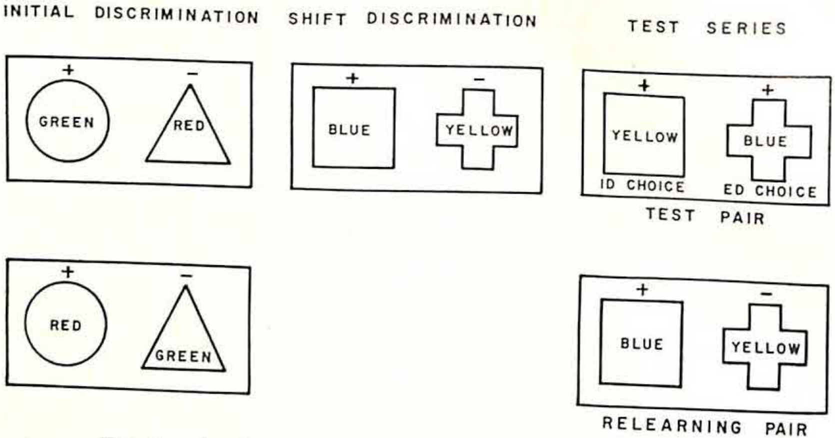


FIG. 1. An illustration of the optional ID-ED paradigm.

replication required 32 *S*s, eight possible (four colors and four forms) positive values in the initial discrimination and four possible positive stimulus compounds during the shift discrimination. Within each age level, each *S* was assigned randomly to 1 of the 32 conditions. The total design was analyzed as a 5 × 2 factorial; 5 age levels and 2 relevant dimensions.

Apparatus.—The apparatus, described in detail elsewhere (Kendler & Kendler, 1970), had a display unit containing two windows through which the stimulus compounds were presented, a start platform, and a well in which marbles could be dispensed. The *S* initiated all trials while response choices were recorded automatically, thus minimizing direct *E–S* interactions after the initial instructions.

Procedure.—The Peabody Picture Vocabulary Test Form A (Dunn, 1959) was first administered to all *S*s except the college students, and then the following instructions were communicated:

> In this game you try to win as many marbles as you can. This is how we play. Do you see these two windows? [*E* points to the two stimulus display windows]. When we are ready they will light up and you will see a design in each window. If you push one of the windows lightly [*E* demonstrates], a marble will come out here [*E* points to marble well]. Which window you press depends on the design. At first you have to guess which design to press but after the first guesses you should learn which design will get a marble every time. Sometimes the design you must press to get a marble will be in this window [*E* points to the left window]. And sometimes it will be in this one [*E* points to the right window]. When this platform lights up, push it [*E* demonstrates] and then use the same hand to choose this one or this one [*E* points to windows].

The preschoolers were assisted through the motion of pushing first the start platform, then one window. On all trials, *S* discriminated between two stimulus compounds that differed both in color and form. Each discriminandum appeared equally often in every block of 12 trials. The number of possible correct successive responses to one position (left or right) did not exceed three. A modified correction procedure was used so that a discriminandum to which an incorrect response was made was repeated on the next trial.

The criterion of learning was 10 successive correct responses. On the initial discrimination, *S* was run to criterion or until the first error after 100 trials, in which case training was terminated and continued on the next day that *S* was available. Interruptions of this sort occurred only during the initial discrimination. Immediately after the criterion was achieved on the initial discrimination, the shift discrimination began. Criterion was again 10 successive correct choices and was followed immediately by the test series. During the test series either choice on the test pair was rewarded, while responses to the relearning pair were rewarded as in the shift discrimination. The *S* was presented with 10 test trials with relearning trials interspersed, the number of which varied from 8 to 12 depending on what part of the program the test series began. If *S* made ID choices on 8 or more of the 10 test trials, he was classified as an ID *S*. An ED *S* made 8 or more ED choices, while failure to make 8 or more consistent responses produced an inconsistent classification, I *S*.

RESULTS AND DISCUSSION

The CA and IQ scores, at each developmental level, are presented in Table 1. A 4 × 2 (Developmental Levels × Relevant Dimensions) analysis of variance of IQ scores revealed no significant differences between ages ($F < 1$), relevant dimensions ($F < 1$), or the interaction between them ($F < 1$). One-way analyses of variance performed at each age level on CA and IQ scores failed to yield any significant

TABLE 1

Chronological Age and Intelligence Quotient Scores for Each Developmental Group

Group	CA (mo.)			IQ		
	M	*SD*	Range	*M*	*SD*	Range
Preschool	54.5	3.1	48–59	113.0	8.7	94–134
Kindergarten	70.7	3.6	65–79	107.4	8.3	89–127
Second grade	94.0	4.2	87–107	111.1	13.8	83–145
Sixth grade	139.9	5.8	133–157	109.9	13.7	90–143
College	229.7	15.2	209–297	—	—	—

differences between color and form groups, thus indicating that within each age level the randomization procedure was successful in equating IQ and CA for the color and form groups.

Table 2 presents the mean number of trials to reach the criterion on the initial discrimination for each of the two relevant dimensions at each age level. The raw scores were subjected to a square-root transformation to correct for unequal variances and then submitted to a 5 × 2 analysis of variance. Significant differences were obtained between ages, F (4, 150) = 37.21, $p < .01$; between dimensions, F (1, 150) = 16.03, $p < .01$; and from the Age × Dimension interaction, F (4, 150) = 2.89, $p < .05$; the last effect resulting from the decreasing difference between color and form scores as age increases. The age component was subjected to a trend test which produced both significant linear and quadratic components: F (1, 150) = 134.19, $p < .01$, and F (1, 150) = 13.07, $p < .01$, respectively, thus indicating both the monotonicity and exponential character of the function.

TABLE 2

Trials to Criterion in the Initial Discrimination as a Function of Age and Relevant Dimension

Group	Color relevant		Form relevant	
	M	*SD*	*M*	*SD*
Preschool	184.6	112.9	98.6	98.9
Kindergarten	71.6	48.8	39.6	40.7
Second grade	28.9	25.1	14.9	13.7
Sixth grade	19.0	22.2	10.4	11.2
College	5.7	5.0	7.6	9.3

TABLE 3

Trials to Criterion in the Shift Discrimination as a Function of Age and Relevant Dimenison in Initial Learning

Group	Color relevant		Form relevant	
	M	*SD*	*M*	*SD*
Preschool	6.5	6.5	7.9	7.0
Kindergarten	3.6	4.6	3.9	6.4
Second grade	1.8	2.3	1.2	1.4
Sixth grade	.6	.5	.3	.3
College	.9	.7	.4	.4

Table 3 reports the mean number of trials to reach the criterion on the shift discrimination for each relevant dimension at each age level. An analysis of variance of the square-root transformed scores yielded only one significant difference, the age variable, F (4, 150) = 8.12, $p < .01$. The age component was subjected to trend tests and again both linear and quadratic components proved significant, F (1, 150) = 26.54, $p < .01$, and F (1, 150) = 5.50, $p < .05$, respectively.

Table 4 reports the proportion of *S*s at each age group, for each relevant dimension, who responded in an ID, ED, or I manner while Fig. 2 graphically represents the combined results. These data were subjected to the L test for linear ranks (Page, 1963) with color and form dimension serving as replications. A significant increasing trend of ID *S*s ($L = 109$, $p < .001$) and significant decreasing trends for ED *S*s ($L = 105.0$, $p < .05$) and I-*S*s ($L = 105.5$, $p < .05$) were found.

Table 5 reports the learning scores of the ID, ED, and I *S*s for the initial and shift

TABLE 4

Proportion of Intradimensional, Extradimensional, and Inconsistent Subjects as a Function of Age and Relevant Dimension in Initial Learning

Group	Color relevant			Form relevant		
	ID	ED	I	ID	ED	I
Preschool	.25	.38	.38	.69	.19	.13
Kindergarten	.50	.31	.19	.75	.13	.13
Second grade	.81	.19	.00	.88	.13	.00
Sixth grade	.88	.13	.00	.81	.19	.00
College	1.00	.00	.00	.94	.06	.00

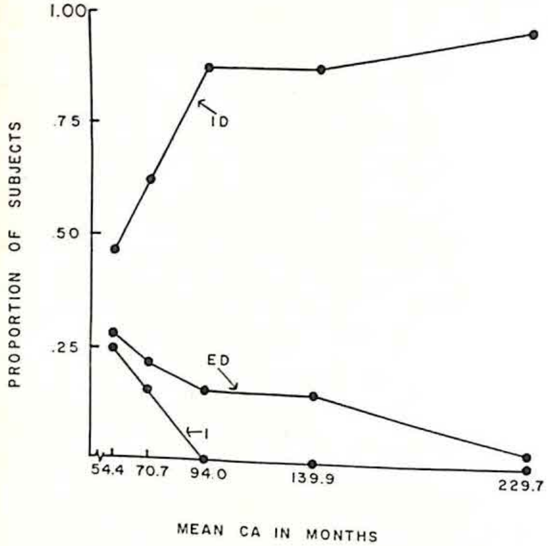


FIG. 2. Proportion of ID, ED, and I *S*s as a function of age.

discriminations. For both training tasks, the ID *S*s learned more rapidly than did the ED *S*s for all five age groups, the chance probability of either occurrence being less than .05. The performance of the I *S*s during the initial discrimination was not significantly different from the ID or ED *S*s. The ID *S*s, however, learned the shift discrimination more rapidly than did the I *S*s both at the preschool and kindergarten levels; $t\ (21) = 1.98,\ p < .05$, and $t\ (23) = 3.27,\ p < .01$, respectively. The differences between ED and I *S*s at these two levels were not significant.

The mean IQs of the ID *S*s at each age level from preschool to sixth grade (114, 109, 112, 110) exceeded those of the ED *S*s (113, 107, 105, 101) at each age level and was also higher than the I *S*s (110, 103) at the two age levels, preschool and kindergarten, at which they were represented. The mean IQ of all ID *S*s (111.1) was significantly different than the mean IQ of ED *S*s (107.5), $t\ (113) = 2.56,\ p < .01$.

Comparison between present results and those of Campione (1970).—A chi-square analysis of preschool and second-grade *S*s classified as ID, ED, and I in this experiment produced a significant difference, $\chi^2\ (2) = 12.5,\ p < .01$. An analysis of variance of the number of ID choices during the test series for these two groups also resulted in a significant age effect, $F\ (1,\ 60) = 6.21,\ p < .05$. Thus those comparisons between preschool and second grade which failed to produce significant differences in Campione's study were found to be significant in the present study. However, neither of the individual comparisons between the same age groups from the two studies proved to be significant below the .05 level. In Campione's study, 63% of the preschool *S*s and 65% of second graders responded in an ID manner, while 45% and 84% *S*s in this study behaved similarly.

One possible reason for the discrepancy between the two studies is the difference in procedure used during the initial discrimination phase. Whereas in the present study all *S*s received the conventional discrimination training procedure, a special technique was used for slow learners in Campione's (1970) study. On the first error following the twentieth trial, *S* was shown all four stimuli, with the pair of positive stimuli separated from the pair of negative stimuli, and was told which of the two pairs of stimuli yielded a reinforcement (a marble) and which pair did not. The slow learners who received special training in Campione's study exhibited the same percentage of ID *S*s as did the fast learners. In the present study, in the absence of any special training techniques, the slow learners at the preschool level, as defined by the criterion Campione used, made significantly fewer ID shifts than did the fast learners; 31% compared to 83%. A similar difference did not occur among

TABLE 5

MEAN NUMBER OF TRIALS TO CRITERION IN THE INITIAL DISCRIMINATION AND THE SHIFT DISCRIMINATION FOR INTRADIMENSIONAL, EXTRADIMENSIONAL, AND INCONSISTENT SUBJECTS

Group	Initial discrimination			Shift discrimination		
	ID *S*s	ED *S*s	I *S*s	ID *S*s	ED *S*s	I *S*s
Preschool	93.1	194.7	163.4	4.7	7.2	11.9
Kindergarten	51.9	73.4	45.8	3.3	4.9	16.4
Second grade	19.7	33.8	—	1.0	4.4	—
Sixth grade	12.2	28.0	—	.4	.6	—
College	6.3	18.0	—	.7	1.0	—

the second graders. Perhaps Campione's special training technique elevated the percentage of ID *S*s among his preschoolers. Consistent with this possibility is the recent finding that displaying the discriminanda simultaneously facilitates reversal learning in kindergartners presumably because it encourages conceptual classification (Kendler & Ward, 1971). The elevation of the percentage of ID *S*s among preschoolers combined with a chance-determined low percentage of ID *S*s among 20 second graders may account for Campione's failure to find a significant difference between his two age groups.

Comparison between optional ID and R shifts.—In both the present and previous study (Kendler & Kendler, 1970) the same number of *S*s and age groups (32 in each of five age groups), apparatus, procedure, and stimulus compounds from the same dimensions were used.

Of major importance is the effect of the two paradigms on the type of shift chosen by *S*s during the test trials. In an analysis of variance of the number of ID or R choices, the only significant result was produced by age, F (4, 310) = 7.97, $p < .01$; the difference between the studies was nonsignificant, $F < 1$. Figure 3 shows the proportion of *S*s in each study that chose the various kinds of optional shifts. Consistent with the apparent similarity of the graphical results are the nonsignificant results of the chi-square analyses of the differences between the frequencies of ID and R, ED, and I choices in the two studies.

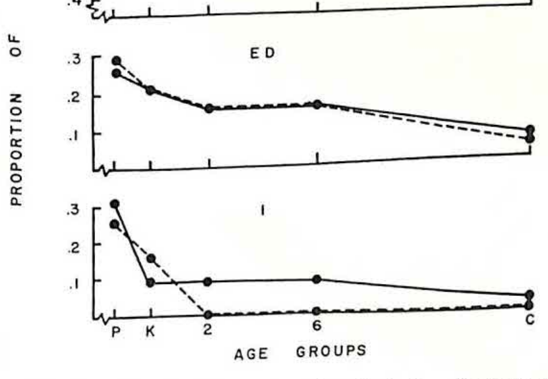


FIG. 3. Comparison of optional choices between the ID–ED, R–ED studies. (The top panel illustrates the proportion of ID or R choices in the two studies, while the middle and lowest panels show the proportion of ED and I choices, respectively.)

The two studies also did not vary in the performance during the initial discrimination. An analysis of variance of the square-root transformation of the number of trials to reach the criterion of learning in the initial discrimination yielded a significant age effect, as expected, F (4, 310) = 52.98, $p < .01$, but no difference between the two studies, $F < 1$. The same analysis performed on the results of the shift discrimination, however, produced both a significant age effect, F (4, 310) = 45.5, $p < .01$, and a significant difference between studies, F (1, 310) = 222.3, $p < .01$. The interaction was also significant, F (4, 310) = 7.9, $p < .01$, resulting from a decreasing difference between studies as age increases. At all age levels, the shift discrimination was learned more slowly in the R–ED procedure than in the ID–ED paradigm; the differences for the lowest four age levels being significant beyond the .01 level, while the difference for the college *S*s was significant beyond the .05 level. One reason for these differences is that in the R–ED procedure the stimulus compounds are unchanged when the reinforcement contingencies are switched from the initial to the shift discrimination thus producing perseverative errors. In contrast, the stimulus compounds in the ID–ED procedure are changed to new dimensional values and as a result half of the *S*s by chance should make a correct response on the first trial of the shift discrimination.

The greater probability of making an error on the first trial of the shift discrimination by the R–ED *S*s, or their tendency to perseverate in making the error, cannot, however, explain entirely the slower learning by these *S*s when compared to the ID–ED *S*s. When the performances of the two groups are compared in the learning of the shift discrimination from the point of making the first correct response, the analysis of variance of transformed scores results in a significant study effect, $F\ (1, 310) = 75.35, p < .01$, age effect, $F\ (4, 310) = 32.80, p < .01$, and interaction effect, $F\ (4, 310) = 5.85, p < .05$. At each age level, except the college *S*s, the R–ED *S*s made significantly more errors, $p < .01$.

CONCLUSION

The present results offer strong support to the notion that ontogenetic changes occur in the tendency for humans to prefer an ID over an ED choice in an optional discrimination-shift paradigm. This empirical generalization applies to the wide age range between $4\frac{1}{2}$ and 21 yr. as well as to the narrow span between $4\frac{1}{2}$ and 8 yr., the two points between which Champione (1970) failed to observe any change. In addition, the results suggest that the ID–ED ontogeny is similar to the R–ED ontogeny (Kendler & Kendler, 1970).

If one assumes that ID choices reflect the operation of a mediational response, then the present data support the conclusion that mediation increases with age. Such a conclusion is at odds with Campione's (1970) assumption that the probability of a child's learning a mediational response remains constant over the age range he tested, and by implication, over the more extended span used in this study. Campione, it should be noted, conceives of mediation as a "response of attending to some dimension of the discriminanda." In contrast the Kendlers postulate the mediational mechanism to be conceptual, not perceptual; the ontogenetic changes observed in the present experiment, as was the case for the similar relationship found with R-ED shifts, reflect the transition from a single-unit continuity mechanism (Kendler, Basden, & Bruckner, 1970; Spence, 1936) to a mediational mechanism with which *S*'s capacity to classify and represent related stimuli is increased.

A mediational interpretation that emphasizes encoding processes gains support from the results of a developmental study in which pictures of conceptual instances (e.g., apple, banana; shirt, pants) were substituted for two-dimensional geometrical patterns (Kendler, Kendler, & Marken, 1969). In spite of the fact that these stimuli failed to provide *S* with an opportunity to attend to any particular dimension or segment of the discriminanda, a developmental change in the case of executing a reversal shift was obtained comparable to those exhibited in optional R-ED studies that used two-dimensional geometrical patterns (Kendler & Kendler, 1970; Kendler et al., 1962) and to the optional ID–ED ontogenetic relationship obtained in the present study.

The present results suggest that the introduction of the new discrimininanda, which distinguishes the ID–ED procedure from the R–ED comparison, does not seem to have much influence on the ontogeny of optional-shift behavior if ID and R choices are considered to be functionally equivalent.

Although the R–ED and ID–ED procedures produce similar results on the test trials, the former generates considerably slower learning during the shift discrimination even when the number of errors during the shift discrimination is limited to those made after the first correct response. The conclusion is that the R–ED procedure produces more negative transfer effects from the initial to the shift discrimination and that these effects are distributed throughout the entire shift-discrimination trials. This finding is not surprising considering the fact that the shift-discrimination task requires *S*s in the R–ED procedure, unlike those in the ID–ED paradigm, to respond to a pair of stimuli in a manner opposite to that which was previously required during the initial discrimination.

The negative transfer effects from the initial discrimination to the shift discrimination are not distributed equally among the three choice groups in either of the two experimental paradigms. The R and ID *S*s exhibit fewer errors than the ED and I *S*s because the former transfer their common mediational representational response from the initial to the shift discrimination. Thus, the pattern of errors during the shift discrimination as well as developmental changes in optional choice responses in both the R–ED and ID–ED paradigms, are consistent with the coordinated single-unit and mediational theory (Kendler & Kendler, 1970).

REFERENCES

CAMPIONE, J. C. Optional intradimensional and extradimensional shifts in children as a function of age. *Journal of Experimental Psychology*, 1970, **84**, 296–300.

DICKERSON, D. J. Performance of preschool children on three discrimination shifts. *Psychonomic Science*, 1966, **4**, 417–418.

DUNN, L. D. *Peabody picture vocabulary test-form B.* Nashville, Tenn.: American Guidance Service, 1959.

KENDLER, H. H. Environmental and cognitive control of behavior. *American Psychologist*, 1971, **26**, 962–973.

KENDLER, T. S., BASDEN, B. H., & BRUCKNER, J. B. Dimensional dominance and continuity theory. *Journal of Experimental Psychology*, 1970, **83**, 309–318.

KENDLER, H. H., & KENDLER, T. S. Vertical and horizontal processes in problem solving. *Psychological Review*, 1962, **69**, 1–15.

KENDLER, H. H., & KENDLER, T. S. Mediation and conceptual behavior. In K. W. Spence & J. T. Spence (Eds.), *Psychology of learning and motivation*. Vol. 2. New York: Academic Press, 1968.

KENDLER, T. S., & KENDLER, H. H. An ontogeny of optional shift behavior. *Child Development*, 1970, **41**, 1–27.

KENDLER, T. S., KENDLER, H. H., & LEARNARD, B. Mediated responses to size and brightness as a function of age. *American Journal of Psychology*, 1962, **75**, 571–586.

KENDLER, H. H., KENDLER, T. S., & MARKEN, R. S. Developmental analysis of reversal and half-reversal shifts. *Developmental Psychology*, 1969, **1**, 318–326.

KENDLER, H. H., KENDLER, T. S., & SANDERS, J. Reversal and partial reversal shifts with verbal material. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 117–127.

KENDLER, T. S., KENDLER, H. H., & WELLS, D. Reversal and nonreversal shifts in nursery school children. *Journal of Comparative and Physiological Psychology*, 1960, **53**, 83–88.

KENDLER, H. H., & WARD, J. W. Single vs. cumulative presentation of stimuli to kindergartners in reversal shift behavior. *Developmental Psychology*, 1971, **5**, 420–426.

KURTZ, K. H. Discrimination of complex stimuli: The relationship of training and test stimuli in transfer of discrimination. *Journal of Experimental Psychology*, 1955, **50**, 283–292.

MUMBAUER, C. C., & ODOM, R. D. Variables affecting the performance of preschool children in intradimensional, reversal, and extradimensional shifts. *Journal of Experimental Psychology*, 1967, **75**, 180–187.

PAGE, E. B. Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Journal of the American Statistical Association*, 1963, **58**, 216–230.

SHEPP, B. E., & Eimas, P. D. Intradimensional and extradimensional shifts in the rat. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 357–361.

SPENCE, K. W. The nature of discrimination learning in animals. *Psychological Review*, 1936, **43**, 427–449.

TIGHE, T. J., & TIGHE, L. S. Discrimination shift performance of children as a function of age and shift procedure. *Journal of Experimental Psychology*, 1967, **74**, 466–470.

ZEAMAN, D., & HOUSE, B. J. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency*. New York: McGraw-Hill, 1963.

(Received November 22, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 110–115

TRANSFER OF IMPLICIT ASSOCIATIVE RESPONSES BETWEEN FREE-RECALL LEARNING AND VERBAL DISCRIMINATION LEARNING TASKS[1]

LAWRENCE E. COLE[2]
Arkansas Polytechnic College

N. JACK KANAK
University of Oklahoma

The present investigation examined the transfer of implicit associative responses (IARs) from a free-recall task to either wrong (W) or right (R) items in a verbal discrimination (VD) task. In the absence of instructions about the relationship between the two tasks, the transfer of IARs to the R items interfered with VD learning, whereas the associative relationship with the W items had no influence upon learning. Informing *S* about the relationship between the R or W items of the VD task and the items of the free-recall task facilitated learning in both conditions. It was postulated that *S*s in the instructed R and W conditions adopted "approach" and "avoidance" strategies, respectively, to the related items.

The present study examined predictions from the frequency theory (Ekstrand, Wallace, & Underwood, 1966) of verbal discrimination (VD) learning concerning the transfer of implicit associative responses (IARs) from a free-recall learning task to a subsequent VD task. Implicit associative responses refer to cognitive activity elicited by highly meaningful verbal units. For example, the presentation of the word MAN is assumed to elicit WOMAN as an IAR and vice versa. These responses are assumed by the frequency theory to add frequency units to the respective items and therefore are expected to enter into determination of the differential frequency cue governing the difficulty of VD learning. However, support for the "transfer" of IARs in single-list VD learning is equivocal (Eberlein & Raskin, 1968; Fulkerson & Kausler, 1969; Kanak, Cole, & Thornton, 1970).

The investigation of IARs has also been extended to a VD transfer task in which the IARs have been manipulated between the wrong (W) or right (R) items of the two lists. In the W_1–R_1, W_2–R_1' paradigm, for example, List 2 R items are associates of the List 1 R items and are paired with new W items. Frequency theory implies that performance in the W_1–R_1, W_2–R_1' paradigm is mediated by the transfer of IARs between the associatively related R items, with a resultant increase in the differential frequency cue, and is therefore dependent upon the application of Rule 1 or selection of the more frequent item. In the W_1–R_1, W_1'–R_2 paradigm, the W items in List 2 are the associates of List 1 W items and are paired with new R items. Under these conditions, Rule 2 or selection of the least frequent item is initially applicable during practice on List 2. With a continuation of practice on the W_1'–R_2 list, Rule 2 should loose its functional significance since R items accumulate frequency units more rapidly than do W items. Thus performance should eventually be dependent upon application of Rule 1. Performance on both paradigms is characteristically compared to a nonspecific transfer control condition which does not include associative relationships between the words in the two lists (i.e., the W_1–R_1, W_2–R_2 paradigm). The frequency theory, however, has not received unequivocal support in studies of VD transfer paradigms.

Kausler and Dean (1967) found that in the absence of instructions regarding the associative relationship between the two lists the positive effect attributed to Rule 2 for the W_1'–R_2 list was demonstrated only on Trial 1. On the subsequent trials performance was inferior to the control. Application of Rule 1 likewise failed to show a pronounced positive effect on the W_2–R_1' list which did not differ significantly from the control list. Two other studies (Kanak & Dean, 1969; Raskin, Boice, Rubel, & Clark, 1968), which also required *S*s to practice without instructions concerning the list relationships, have reported similar findings. Half of the *S*s in the Raskin et al. study were informed about the relationship between the lists, and, as predicted by frequency theory, performance on the W_2–R_1' list was superior to the W_2–R_2 list. Performance on the W_1'–R_2 list was inferior to the control list on all trials, thus providing evidence for mediated associative interference. Consequently, these authors noted the importance of instructions in producing IAR activity in transfer tasks.

---

[1] The authors wish to acknowledge the assistance of Charles Sikes, who collected the data.

[2] Requests for reprints should be sent to Lawrence E. Cole, Department of Behavioral Sciences, Arkansas Polytechnic College, Russellville, Arkansas 72801.

The studies cited above imply that the positive increment to the differential frequency cue deriving from the exchange of IARs among associatively related items in a VD transfer task is substantially mitigated by incidental associative competition mechanisms when *S*s are uninformed about the interlist relationship between W or R items. The present study attempted to enhance IAR activation by replacing the first VD list with a free-recall learning task which contained associates of subsequent VD-list W or R items. The free-recall learning procedure ensures the priming of implicit associative response mechanisms but avoids the development of intrapair incidental associative learning (Sardello & Kausler, 1967) and the resultant interference in transfer associated with prior List 1 VD learning. Others (e.g., Kanak & Dean, 1969; Kausler & Dean, 1967) have shown that incidental associative interference may be involved in the W_1–R_1, W_2–R_1' and W_1–R_1, W_1'–R_2 paradigms as well as other paradigms involving potential associative competition—unlearning relationships. In addition, in further support of the notion of associative interference, Kausler, Fulkerson, and Eschenbrenner (1967) have demonstrated that List 1 incidental associations undergo unlearning in the verbal discrimination analogue (i.e., W_1–R_1, W_1–R_2) of the A–B, A–C paired-associate transfer paradigm. The present prior free-recall learning task prevents intrapair associations of the type involved in the above studies from being learned. As a result, IAR mechanisms may be activated without the interfering effect of incidentally learned associations if such activation were accomplished by a prior VD list. Performance on the VD task should therefore only be influenced by frequency mechanisms, or by whatever alternative strategies *S* may employ, since no available evidence suggests that associative mechanisms affect the proficiency of learning a single standard VD list.

Following free-recall learning, one-third of the *S*s learned a 12-pair VD list in which the 12 R items were the highest normative associates of the 12 items involved in the free-recall task and were paired with 12 neutral W items (Cond. TR) which did not appear in the free-recall task. The 12 neutral items, however, had normative primary associates which were approximately equivalent in associative frequency to the 12 items of the free-recall task, although these associates were not utilized. As can be determined in Table 1, this procedure insured that *S*s could not select the R item in the VD task on the basis of a difference in meaningfulness, compared to the W item. Similarly, one-third of the *S*s learned a subsequent VD list, after free-recall learning, in which the 12 W items were the same highest normative associates of the 12 free-recall task items and were paired with the 12 neutral R items (Cond. TW). Within each of the above conditions, one-half of the *S*s were instructed (I) regarding the associative relationships between R or W words of the two tasks while the remainder were uninstructed (U). The findings of Raskin et al. (1968) suggest that the positive contributions of IARs should be confined to the instructed conditions. The remaining one-third of the *S*s constituted a control condition (C) in which one-half of the *S*s learned the same VD list as the TR conditions and the other one-half learned the same list as the TW conditions, but the free-recall learning task involved items unrelated to either of these VD lists.

It should be apparent that the TR and TW conditions are analogous to the relationships involved in the W_1–R_1, W_2–R_1', and W_1–R_1, W_1'–R_2 transfer paradigms, respectively, with regard to the intertask manipulation of IARs, but lack the opportunity for prior intrapair associative learning. According to frequency theory, therefore, *S*s in the TR condition should respond during VD learning by selecting the more frequent item (Rule 1), while Rule 2 or selection of the least frequent item is initially applicable in the TW condition. Later in learning, however, Rule 2 should become inappropriate in the TW condition as a consequence of the breakdown in the frequency differential associated with the more rapid increment in frequency units for R items. It should also be noted that the TR and TW conditions are also similar to the associate right (AR) and associate wrong (AW) conditions investigated by Kanak et al. (1970). These authors reported little support for the frequency theory, and the present study therefore represents a further test of AR and AW interitem manipulations under conditions more favorable to the active elicitation of IARs.

TABLE 1

ACQUISITION LISTS FOR FREE-RECALL AND VERBAL DISCRIMINATION LEARNING

Free-recall learning items		Verbal discrimination pairs			
Experimental	Control	Original		Re-paired	
DOGS	NEEDLE	SHOES	NEVER[a]	SPIDER	QUEEN[a]
TABLE	HAND	COLD	FLOWERS	HUNGRY	SOFT
YOUNGER	BITTER	HIGH	WHITE	ANGER	WHITE
HARD	TOWN	OCEAN	CATS	HIGH	OLDER
SLOW	ROUND	AFRAID	SOFT	PEPPER	BIRD
ALWAYS	DOCTOR	LAMP	QUEEN	OCEAN	CHAIR
BLACK	BREAD	HUNGRY	CHAIR	BUTTER	CATS
LONG	TRUCKS	BUTTER	SHORT	AFRAID	SHORT
EAGLE	NAIL	SPIDER	OLDER	CARPET	THERE
HERE	WINDOW	CARPET	FAST	LAMP	NEVER
BLOSSOM	LOST	ANGER	BIRD	COLD	FAST
KING	HOME	PEPPER	THERE	SHOES	FLOWERS

[a] Associates of free-recall task items are in the right spatial position.

METHOD

Subjects.—Eighty volunteers from the undergraduate classes at Arkansas Polytechnic College participated in the experiment. All *S*s were naive with respect to prior participation in VD learning research. Placement of *S*s into the respective treatment conditions was made according to a predetermined random block assignment.

Materials.—Twelve pairs of associates (DOGS-CATS, TABLE-CHAIR, YOUNGER-OLDER, HARD-SOFT, SLOW-FAST, ALWAYS-NEVER, BLACK-WHITE, LONG-SHORT, EAGLE-BIRD, HERE-THERE, BLOSSOM-FLOWERS, and KING-QUEEN) were selected from the Palermo and Jenkins (1964) word association norms. From this pool, the 12 primary stimulus words served in the free-recall learning task. Their associates were randomly paired in the VD task with a new set of 12 unrelated words also selected from the Palermo and Jenkins norms. For the TR conditions, the associates of the free-recall items served as R items in the VD task and the W-R item function was simply reversed for the TW conditions. The C condition practiced on the same VD list as the other conditions, but their 12 free-recall task items were words unrelated to those in the VD list and were also selected from the Palermo and Jenkins' norms. The lists used in the free-recall learning task and the VD task are presented in Table 1.

To control for idiosyncratic associations and to increase generalizability of the findings, for one-half of the *S*s the resulting VD lists were randomly re-paired such that original pairings were not retained. The standard procedures of randomizing serial and spatial position were also completed for each VD list variant. The free-recall learning tasks were also presented in four random orders to minimize serial learning effects.

Procedure.—Each *S* was informed that alternation of study and recall trials, characteristic of the free-recall task, would continue until he was able to recall all 12 words. In the study trials, the 12 words were presented individually for 2 sec. by means of a tape recorder. After each study trial, *S* was given a sheet of paper on which 12 lines were placed in vertical alignment and was asked to recall as many words as possible in any order. Forty-five seconds were allowed for each recall trial. Following the free-recall trials, instructions standard to the anticipation procedure of VD learning were administered to each *S*. In addition, *S*s in the I-TR and I-TW conditions were informed about the nature of the relationship between the free-recall learning items and the VD pairs. The *S*s in the other conditions were simply told the free-recall task was an integral part of the experiment. The VD pairs were then administered via the tape recorder by the anticipation procedure at 2:2 sec. rate. Practice on the VD list was carried to a criterion of 3 successive perfect trials or 10 trials, whichever came first. Only 10 *S*s required 10 trials to reach the alternate criterion, 5 in the U-TR condition, 2 in the C condition, and 1 each in the U-TW, I-TW, and I-TR conditions.

Upon completion of practice on the VD list, a terminal free-recall task was administered without previously informing *S*. This task consisted of each

S recording as many W and R items as he could remember without regard to item function or order on a sheet of paper which presented 24 lines in vertical alignment. The terminal free-recall task was employed to measure the availability of W and R items for recall. To the extent that IARs are activated in the experimental conditions, W-item recall should be superior to the control condition in the TW conditions and R-item recall should be superior in the TR conditions. Furthermore, if instructions facilitate the activation of IARs, these effects should be most pronounced in the I-TW and I-TR conditions.

RESULTS

Preliminary analyses showed that the variants of both the free-recall list and the VD list were comparable across the treatment conditions, *F*s < 1.

Variability in the difficulty of learning the VD list, as reflected in the total number of errors to criterion, was highly significant, F (4, 75) = 13.86, $p < .005$. The means and standard deviations ($n = 16$) are presented in Table 2. The Newman-Keuls test, used to make the pairwise comparisons, revealed that the U-TR and I-TR conditions differed significantly from the remaining conditions and from each other (all ps $< .05$). Regarding the effect of instructions, both the I-TR and I-TW conditions resulted in significantly fewer errors than the C condition (ps $< .01$ and $< .05$, respectively), the U-TR condition was more difficult than the C condition ($p < .05$), while the U-TW condition did not significantly differ from the C condition ($p > .05$).

Recall of W and R items was analyzed via a 5 (Treatment Conditions) × 2 (W and R items) analysis of variance with the last factor treated as a within-*S* factor. Significant sources of variability included item recall, F (1, 75) = 150.02, $p < .001$, and the interaction, F (4, 75) = 21.06, $p < .005$. Treatment conditions produced a trend toward significance, F (4, 75) = 2.35, $p < .10$. The cell means and standard deviations for W-item and R-item recall are summarized in Table 2. Statistical comparison of the means for W-item recall showed that the I-TW conditions resulted in more W items recalled than each of the other conditions (ps $< .01$). The superiority of the U-TR and C conditions to the I-TR condition represented the only other significant comparisons (ps $< .05$). For R-item recall, the I-TR condition produced more items recalled than the other conditions, but it was significantly superior to only the I-TW condition ($p < .05$) which had the least number of R items recalled. Since differences in rate of learning could be correlated with the recall scores, product-moment correlations were computed between trials to criterion and total recall. The only significant correlation was for the I-TR condition, $r\ (14) = -.44$, $p < .05$.

TABLE 2

MEANS AND STANDARD DEVIATIONS FOR LIST 2 ERRORS AND W-ITEM AND R-ITEM TERMINAL FREE RECALL

Cond.	List 2 errors		W-item recall		R-item recall	
	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*
U-TR	14.94	6.34	5.13	1.74	7.75	2.20
C	10.88	4.11	5.00	2.22	8.31	1.40
U-TW	9.69	4.43	8.06	2.32	7.81	1.76
I-TW	6.75	4.28	4.38	2.50	7.31	1.74
I-TR	3.44	3.65	3.19	1.22	9.38	2.06

In agreement with Raskin et al. (1968), the present results demonstrate the importance of informing *S* about any associative relationship involving the List 1 and List 2 W and R items of a transfer paradigm. Frequency theory assumes that IARs do transfer from List 1 to List 2 in transfer tasks involving associatively related items and that, therefore, performance in the W_2–R_1' and W_1'–R_2 second lists should be regulated by Rules 1 and 2, respectively. It is further implied that performance on a W_1'–R_2 list will eventually depend upon the application of Rule 1 since frequency units rapidly accrue to the R items during learning. In a similar manner, IARs should accrue to the associates of the free-recall learning items while *S* practices on the free-recall task. In the TR conditions, the transfer of IARs should add frequency units to the R items and frequency units should be added to the W items in the TW conditions with or without instructions concerning the intertask relationships. The appropriate Rules should then be 1 and 2, respectively, with Rule 1 later applying to the TW conditions in the later stages of learning.

The present findings, however, do not fully support previous results reported for the W_2–R_1' and W_1'–R_2 paradigms nor do they well accord with frequency theory. Without instructions, *S*s in the U-TR condition tended to select the unrelated or presumably the least frequent item rather than the most frequent, whereas the U-TW condition failed to differ significantly from the C condition in contrast to its predicted inferiority. Performance on Trials 1–4 was also examined for all conditions, but the results replicated the analysis on total errors. Previous investigations concerned with the W_2–R_1' and W_1'–R_2 paradigms (Kanak & Dean, 1969; Kausler & Dean, 1967; Raskin et al., 1968) did not report significant differences, in the absence of instructions, but found the W_2–R_1' list somewhat easier than the W_2–R_2 or control list. Performance on the W_1'–R_2 list was essentially inferior to the W_2–R_2 condition in the early trials, a result attributed to mediated associative interference (Kausler & Dean, 1967). In the present study, the comparable U-TW condition tended to be easier than the C condition, in the absence of an associative interference component.

It is interesting to note the similarity between the present findings and those reported by Kanak et al. (1970), who manipulated IARs between W and between R items in a single VD list. Contrary to frequency theory, manipulation of IARs in the AW condition failed to have an inhibitory effect upon performance, and, instead of the facilitation predicted by frequency theory, these authors found that the AR condition was the most difficult. The pronounced retardation demonstrated when IARs were manipulated in the U-TR condition of the present study and in the AR condition of the Kanak et al. study is perhaps attributable to interference created by extraexperimental words elicited as IARs during practice on the VD list. Barch, Lippman, and Whalen (1967) have suggested that these alternate associative links are a part of the adult's language habits. As the result, these can be elicited as IARs during learning and adversely affect the learning process.

Instructions apparently inhibited the interference from extraexperimental IARs, or activated other learning strategies, because, in sharp contrast to their uninstructed counterparts, both the I-TR and I-TW conditions were superior to the C condition. Seemingly, *S*s successfully applied Rule 1 and Rule 2, respectively. However, *S*s in the U-TW condition did not experience the interference in later learning expected as a result of the rapid accumulation of frequency units to the R items and the consequent predicted necessity of changing to a Rule 1 strategy. Thus, perhaps *S*s in the instructed conditions employed some other strategy. Recent research concerning reversal shifts in VD transfer tasks (Paul & Paul, 1968) and intrapair manipulations of IARs (Barch et al., 1967) have shown that *S*s can apply functional rules other than selecting R items on the basis of relative frequency. Therefore, instead of applying rules based upon frequency, *S*s in the I-TR and I-TW conditions may have adopted the rules "approach the related items" and "avoid the related items," respectively. Furthermore, it is logical to assume that if *S*s in the I-TW condition actively avoided the W items and *S*s approached the R items in the I-TR condition, the act of utilizing these rules would no doubt enhance the learning of these items as "responses." The predominance of W and R items in the terminal free-recall protocols of the I-TW and I-TR conditions, respectively, gives credence to this hypothesis.

In view of the present findings, it appears that interpretations of VD learning involving manipulations of IARs need to be expanded to include explanations other than frequency theory. The results obtained in the instructed conditions of the present experiment, as well as those reported by Barch et al. (1967) and Paul and Paul (1968), suggest that *S* will adopt any strategy of VD learning which he discovers to be more efficient.

REFERENCES

BARCH, A. M., LIPPMAN, M., & WHALEN, S. Intrapair relations in verbal discrimination learning: A paradoxical finding. Paper presented at the meeting of the Midwestern Psychological Association, Chicago 1967.

EBERLEIN, E., & RASKIN, D. Intrapair and interpair associations in verbal discrimination learning. *Psychonomic Science*, 1968, **11**, 145–146.

EKSTRAND, B. R., WALLACE, W. P., & UNDERWOOD, B. J. A frequency theory of verbal-discrimination learning. *Psychological Review*, 1966, **6**, 566–578.

FULKERSON, F. E., & KAUSLER, D. H. Effects of intrapair and interpair bidirectional associates on verbal-discrimination learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 307–310.

KANAK, N. J., COLE, L. E., & THORNTON, J. W. Inter-item manipulations of implicit associative responses in verbal-discrimination learning. *Psychonomic Science*, 1970, **18**, 359–361.

KANAK, N. J., & DEAN, M. F. Transfer mechanisms in verbal discrimination. *Journal of Experimental Psychology*, 1969, **79**, 300–307.

KAUSLER, D. H., & DEAN, M. O. Direct and mediated transfer for right and wrong responses in verbal discrimination learning. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 672–674.

KAUSLER, D. H., FULKERSON, F. E., & ESCHENBRENNER, A. J., JR. Unlearning of List 1 right items in verbal-discrimination transfer. *Journal of Experimental Psychology*, 1967, **75**, 379–385.

PAUL, C., & PAUL, H. Transfer-activated response sets in verbal learning and transfer. *Psychological Review*, 1968, **75**, 537–549.

PALERMO, D. S., & JENKINS, J. J. *Word association norms: Grade school through college.* Minneapolis: University of Minnesota Press, 1964.

RASKIN, D. C., BOICE, C., RUBEL, E. W., & CLARK, D. Transfer tests of the frequency theory of verbal discrimination learning. *Journal of Experimental Psychology*, 1968, **76**, 521–530.

SARDELLO, R. J., & KAUSLER, D. H. Associative recall in verbal-discrimination learning as related to pronunciation and degree of practice. *Psychonomic Science*, 1967, **8**, 253–254.

(Received November 24, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 116–122

ACQUISITION AND EXTINCTION RATES AS DETERMINANTS OF AGE CHANGES IN DISCRIMINATION SHIFT BEHAVIOR[1]

DONALD J. DICKERSON,[2] NEIL NOVIK, AND SHARON A. GOULD

University of Connecticut

The learning of kindergarten and second-grade children was compared under two conditions of discrimination transfer. The experimental procedure for each *S* involved the learning of three consecutive discrimination problems. The same stimuli, which varied in size and brightness, were employed in all problems. One dimension was relevant and the other was irrelevant and variable within trials on each problem. Problem 2 was an extradimensional shift from Problem 1, and Problem 3 was another extradimensional shift from Problem 2. Problem 3 was either the same as (Same condition) or a reversal of (Reversal condition) Problem 1. Kindergarteners learned faster under the Same than under the Reversal condition. Condition differences were absent for second graders. The results suggest that kindergarten children extinguish mediating responses faster than instrumental choice responses, while the reverse probably holds with second graders.

Interest in the development of human discrimination learning was engendered by Kendler and Kendler (1962) when they suggested that the manner in which discrimination problems are learned changes ontogenetically. Specifically, these writers proposed that children below the age of approximately 5 yr. learn in a single-link, S–R fashion, while older humans learn in a covert, mediational fashion. In the single-link type of learning (e.g., Spence, 1936), differential response tendencies are developed directly to the discriminative cues through the effects of reward and nonreward. The *S* is viewed as responding to the stimulus with the greater response tendency. In the mediational type of learning, *S* acquires a chain of two responses, an implicit mediating response to the relevant stimulus dimension and an instrumental choice response to the rewarded cue on that dimension. The function of the mediating response, which may have either a perceptual or a verbal basis, is to alter the stimulus input in a manner such that the effective discriminative cues become the values of a specific dimension.

Kendler and Kendler (1962) supported their developmental change hypothesis with data from two types of experiments. One type consisted of comparing the ease of learning discrimination reversals and extradimensional (ED) shifts (in which a previously irrelevant dimension is made relevant). Single-link learners (young children) should find reversal shifts, which involve negative cue transfer, more difficult than ED shifts, which generally involve no consistent source of cue transfer. Alternatively, mediators (older humans) should find ED shifts more difficult than reversal shifts since these situations involve, respectively, negative and positive transfer of mediating responses. The optional shift procedure is employed in the second type of experiment. In these experiments, *S* first is trained to criterion on a problem involving a relevant dimension (e.g., size) and an irrelevant dimension (e.g., brightness) which varies within trials. The two dimensions are made redundant and relevant in a second phase, with the positive complex containing the previously negative cue. The *S* can learn this problem either as a reversal or as an ED shift. The

---

[1] This investigation was supported by National Institute of Mental Health Training Grant HD-00084-06 from the Department of Health, Education, and Welfare to the University of Connecticut. Preparation of the final manuscript was partially supported by United States Public Health Service Training Grant HD-00183 from the National Institute of Child Health and Human Development to the Kansas Center for Research in Mental Retardation and Human Development.

[2] Requests for reprints should be sent to Donald J. Dickerson, Department of Psychology, University of Connecticut, Storrs, Connecticut 06268.

final phase of the procedure is a transfer test which allows an assessment of the way in which the optional shift was learned. The prediction from the developmental change position is that the proportion of *S*s making reversal shifts should increase with age since such a choice is consistent with the mediational type of learning.

The developmental change hypothesis has not been consistently supported by data. In experiments comparing the ease of learning reversal and ED shifts, adult humans perform in a manner consistent with the mediational position (Isaacs & Duncan, 1962; Kendler & D'Amato, 1955). But so, apparently, do young children when appropriate procedures are used (for a summary of the pertinent results see Caron, 1970). The results of optional shift studies clearly indicate an increase with age in the proportion of *S*s making reversal shifts (Kendler & Kendler, 1968; Kendler, Kendler, & Learnard, 1962). In this instance, however, what is not clear is that the observed trend is due to a change from single-link to mediational functioning. Tighe and Tighe (1966) reported that almost all of their 3- and 4-yr.-old *S*s made reversal shifts when trained initially on a very salient dimension. More recently, Campione (1970) found that if preschool and second-grade children were allowed an option of making either intradimensional (in which new cues preclude a reversal) or ED shifts, approximately two-thirds of both groups showed the former behavior.

The foregoing discussion indicates that the hypothesis of an ontogenetic change from single-link to mediational functioning is questionable at best. More tenable is the position that young children, as well as older children and adults, learn in a mediational fashion. The data, particularly those obtained with the optional shift procedure, do suggest, however, that some type of change is taking place developmentally, and a different interpretation has been offered for these data (Campione, 1970; Dickerson, Wagner, & Campione, 1970). This interpretation has two main points: (*a*) Mediational learning is characteristic of all ages, and (*b*) the relative rates at which mediating and instrumental choice responses are acquired and extinguished change developmentally. Specifically, it was proposed that young children acquire and extinguish choice responses slower than mediating responses and that this difference is reversed in older children and adults.

The present experiment was designed to test the foregoing interpretation. The procedure was modeled after one used by Lawrence and his associates (Goodwin & Lawrence, 1955; Lawrence & Mason, 1955) and involved the learning of three consecutive problems by two different groups of *S*s. The same stimuli were used in all three problems, and these stimuli varied along the dimensions of size and brightness. For the first two problems, the two groups were treated identically. The *S* learned the first problem, which involved a relevant (e.g., size) and an irrelevant (e.g., brightness) dimension, and then learned the second problem, which was an ED shift from the first problem. The third problem was another ED shift, with the relevant dimension of the first problem becoming relevant again. For one group the solution of the third problem was the same as the solution of the first problem, and for the other group the third problem was a reversal of the first problem. The *S* theoretically will begin the second problem with a strong irrelevant mediating response and a strong choice response to a cue on this dimension. Both responses initially will occur in conjunction and will be rewarded and nonrewarded on a chance basis. Both prepotent responses thus will undergo extinction. The irrelevant mediating response will at some point be extinguished and *S* will acquire a mediating response to the relevant dimension. When this occurs, the choice response to the previously positive cue can either be fully extinguished or not. If it is not, this choice response theoretically cannot undergo further extinction while *S* is making a mediating response to the other dimension, i.e., during the remainder of the second problem. (This last statement is based upon the assumption made by most mediational theories that changes in the tendency to select a particular cue occur only when a mediating response is made to the appropriate dimension.) During the learning of the third problem, *S* again will change his mediating response, this time back to the dimension which also was relevant in the first problem. If his choice response to the previously positive cue was not fully extinguished during the learning of the second problem, then he will tend to choose this cue again. Faster learning should result if this cue is positive (Same condition) than if it is negative (Reversal condition). If the choice response to the previously positive cue was fully extinguished during the second problem, then the two conditions should yield similar learning rates.

The performance of kindergarten and second-grade children under these conditions was compared in the present experiment. Kendler et al. (1962) reported that only 50% of their kindergarten sample made optional reversal shifts. This result, when interpreted from the present viewpoint, suggests that kindergarten children may acquire and extinguish choice responses slower than mediating responses. They therefore should learn faster when the third problem is the same as the first problem than when the third problem is a reversal of the first problem. Second graders supposedly acquire and extinguish choice responses faster than mediating responses and, therefore, should show similar rates of learning under both conditions.

METHOD

Subjects.—The *S*s were 16 male and 16 female kindergarten children and 16 male and 16 female second-grade children, all enrolled in a public school in Manchester, Connecticut. Age ranged from 64 mo. to 78 mo. in the kindergarten sample, with a mean of 69.5 mo., and from 89 mo. to 112 mo. in the second-grade sample, with a mean of 95.8 mo. The children may be described as coming from predominantly white, middle-class, and Protestant backgrounds. Each *S* was assigned randomly to one of four conditions, with the restriction that an equal number of male and female *S*s was assigned to each condition.

Apparatus.—A portable discrimination learning apparatus was used. The apparatus consisted of a flat base, 25.4 cm. wide × 48.3 cm. long, and a movable cover, 50.8 cm. long and 35.6 cm. high. Centered 14.0 cm. apart in the base were two circular reward wells, 6.4 cm. in diameter and 6.3 mm. deep. The movable cover was attached to the base in a manner such that when lowered it blocked *S*'s view of *E*'s manipulations of the stimuli, and when raised it allowed *S* to view and to respond to the stimuli. The apparatus was painted dark brown, and the bottoms of the reward wells were lined with green felt.

The stimuli were four wooden cylinders, 3.2 cm. in diameter, which varied in height, 15.2 cm. versus 8.9 cm., and in brightness, black versus white. Each stimulus was mounted on a 8.9-cm.-sq. masonite base which was painted a medium gray. Two such bases with no stimuli attached were used during pretraining.

General procedure.—The experimental treatment for each *S* required a minimum of three and a maximum of six daily sessions. Each *S* was trained twice to a criterion of nine correct responses in 10 consecutive trials on each of three discrimination problems. The stimuli described above were employed on all problems.

The *S*'s response was pushing back a stimulus to expose the reward well under it. Red and blue plastic chips, unsystematically varied, served as rewards. The *S* deposited these chips as they were acquired into a gray metal can with a small opening in the top. In addition to the chip, correct responses were rewarded with "good." Incorrect responses were punished with "no," and correction was not permitted.

Pretraining.—A brief period of pretraining preceded training on the first problem. The *E* first told *S* that he "was going to play the chip game," and that *S* was to attempt to find the chip by pushing back one of the stimuli each time *E* raised the cover of the apparatus. Three practice trials with the plain stimulus bases then were given. Both reward wells contained chips on these trials. On the first trial, the bases were placed in back of the uncovered reward wells, and the chips clearly were visible. The reward wells were only partially covered on the second trial and were fully covered on the third trial. Next, *S* was shown the four stimuli and was asked to point in a predetermined order to the pairs having each of the different cue values, i.e., short, tall, black, and white. Eight more pretraining trials were then given. On these trials, the experimental stimuli were employed and varied on each trial along both the size and brightness dimensions. The *S* was told prior to each response which of the four cues was associated with the chip reward. Each cue was correct for two consecutive trials and was paired once on these trials with each value of the other dimension and with each position. The order in which a given *S* was asked to point to the pairs of stimuli with specific cue values and in which the different cues were correct on the subsequent pairs of practice trials were the same. The *S* pointed and responded to cues in the following order: (*a*) a randomly selected cue on the irrelevant dimension of the first problem, (*b*) the positive cue of the first problem, (*c*) the other cue on the irrelevant dimension, and (*d*) the negative cue of the first problem.

Training.—The *S* began training on Problem 1 immediately upon the completion of pretraining. The *E* introduced this problem by stating that it was now *S*'s turn to guess which stimulus had the chip under it and that *S* should attempt to find the chip every time.

The training procedure was essentially the same for all three problems of the sequence. A training session was terminated either (*a*) by the attainment of criterion performance on a problem or (*b*) by the first error after the fortieth trial which eliminated the possibility that *S* was in a criterion run. Training was continued for 1 more day (40 trials) on a problem if criterion was not reached in the first session of training. On the day following criterion performance for Problems 1 and 2, *S* first was retrained to criterion, and the next problem in the sequence then was introduced immediately without comment from *E*. After every tenth precriterion trial on all problems, *E* said to *S*, "Remember, the game is to find the chip every time."

Any *S* who failed to reach criterion in two sessions on either Problem 1 or Problem 2 was dropped from the experiment and replaced with another *S* from the available pool. Twelve kindergarteners and four second graders were replaced.

Conditions.—One of the dimensions, size or brightness, was relevant and the other was irrelevant and varied within trials on each problem. For half of the *S*s at each age level, the relevant dimensions were brightness in Problem 1, size in Problem 2, and brightness in Problem 3. For the other half of the *S*s, the order of relevant dimensions was size, brightness, and size. Each of the Age Level × Sequence groups was divided once again. The solution of Problem 3 was the same as that of Problem 1 for one of these subgroups (Same condition), while the solution of Problem 3 was a reversal of that of Problem 1 for the other subgroup (Reversal condition). For example, an *S* in the Same condition might have white, tall, and white as positive cues in Problems 1, 2, and 3, respectively. A similar *S* in the Reversal condition would have black as the positive cue in Problem 3.

The positive cues on the relevant dimensions in Problems 1 and 2 were determined randomly for each *S*. The position of the correct stimulus, left or right, over trials was varied according to a Gellermann series. Each irrelevant cue was paired with the positive cue on 5 of every 10 trials. This pairing was restricted so that no more than 3 consecutive trials were alike.

RESULTS

The principal measure employed was the number of errors through criterion. The means and standard deviations of the error scores are presented in Table 1 for all subgroups, disregarding the sex variable. The data for each problem were analyzed in a 2 × 2 × 2 × 2 factorial design, with sex (male vs. female), age level (kindergarten vs. second grade), relevant dimension sequence (brightness-size-brightness vs. size-brightness-size), and condition (same vs. reversal) as the factors.

Only the effect of relevant dimension sequence attained significance in the analy-

TABLE 1

MEANS AND STANDARD DEVIATIONS OF ERRORS-THROUGH-CRITERION SCORES

Level	Cond.	Dimension sequence	Measure	Problem 1	Problem 2	Problem 3
Kindergarten	Same	SBS	$\bar{X}$	3.50	15.12	5.00
			SD	4.50	9.88	3.85
	Same	BSB	$\bar{X}$	19.75	12.50	4.50
			SD	17.40	6.63	3.12
	Reversal	SBS	$\bar{X}$	3.50	12.50	20.87
			SD	2.78	6.87	12.43
	Reversal	BSB	$\bar{X}$	12.50	11.75	9.62
			SD	8.54	9.51	6.59
Second grade	Same	SBS	$\bar{X}$	8.63	11.75	5.75
			SD	8.55	7.78	3.73
	Same	BSB	$\bar{X}$	8.00	7.12	5.75
			SD	10.11	1.36	4.43
	Reversal	SBS	$\bar{X}$	3.87	19.12	4.13
			SD	4.12	12.73	3.09
	Reversal	BSB	$\bar{X}$	12.75	5.37	5.25
			SD	10.40	2.97	6.52

Note.—SBS = size-brightness-size. BSB = brightness-size-brightness.

ses of the data from Problem 1, F (1, 48) = 11.32, and Problem 2, F (1, 48) = 7.48, both $ps < .01$. As can be seen in Table 1, these effects seem to have resulted from generally faster learning when size was the relevant dimension of either problem.

Two main effects and two interactions attained statistical significance in the analysis of the Problem 3 data. Primary interest is focused upon the significant interaction of the age and condition factors, F (1, 48) = 13.28, $p < .001$, since faster learning under the Same condition than under the Reversal condition was expected only for the kindergarten *S*s. In order to assess whether the significant interaction reflected the expected pattern of results, F tests were made between the two conditions at each age level. Learning was faster for kindergarteners under the Same conditions than under the Reversal condition, F (1, 24) = 14.38, $p < .001$. The conditions effect was nonsignificant for second graders, F (1, 24) = .47. The significant main effects obtained for age, F (1, 48) = 9.09, and condition, F (1, 48) = 8.85, both $ps < .01$, are of little importance given the significant interaction of the two factors.

The other significant interaction in the analysis involved the age and relevant dimension sequence factors, F (1, 48) = 4.12, $p < .05$. Subsequent comparisons showed that Problem 3 was learned faster by kindergarteners when brightness was relevant than when size was relevant, F (1, 24) = 4.50, $p < .05$, while the dimension effect was nonsignificant at the second-grade level, F (1, 24) = .13. The dimension effect obtained for kindergarteners is the reverse of the overall effects found in Problems 1 and 2.

The error data were analyzed in another fashion in order to correct for individual differences in speed of learning. The error scores of Problem 1 were subtracted from those of Problem 3 for each *S*, and the difference scores then were analyzed in the same manner as the original error scores. The analysis yielded essentially the same results as the analysis of the Problem 3 error scores. The interaction of age and condition was significant, F (1, 48) = 9.25, $p < .01$. This interaction was due to a conditions effect for kindergarteners, F (1, 24) = 11.21, $p < .01$, and the absence of such an effect for second graders, F (1, 24) = .16.

An analysis also was made of perseverative responding at the outset of Problems 2 and 3. The present interpretation implies that kindergarten children extinguish choice responses slower than second-grade children. If so, then one might expect the kindergarteners of the present experiment to show a longer run of consecutive responses to the previously positive, and now irrelevant, cue upon being shifted from Problem 1 to Problem 2 and from Problem 2 to Problem 3. The number of errors made during the perseverative runs on Problem 2 and Problem 3 were tabulated for each *S* and were analyzed in a five-factor (the four factors indicated above plus that of problems) analysis of variance. The only significant effect in this analysis was that of age, F (1, 48) = 6.61 $p < .05$. The kindergarteners and second graders showed 1.81 and 1.31 mean perseverative errors per problem, respectively.

Also studied was the nature of the response made by each *S* on the trial immediately following the last perseverative response on Problem 3. If kindergarten children extinguish choice responses slower than mediating responses, then a kindergarten *S* should tend to terminate the run of perseverative responses occurring at the start of Problem 3 with the extinction of the mediating response and not of the instrumental choice response. The *S* then should tend on the next trial to make a mediating response to the relevant dimension of Problem 1. If so, the present view predicts a choice response to the positive cue of Problem 1. In the case of second-grade children, the run of perseverative responses should tend to terminate with the extinction of the choice response and not of the mediating response. The choice response thus should tend to be to the other cue on a previously relevant dimension. A count was made of those *S*s who responded on the trial on which the perseverative run was broken on Problem 3 to the complex containing the positive cue of Problem 1. Twenty-four of 32 (75%) kindergarten *S*s and 12 of 32 (37.5%) second-grade *S*s made such responses. Assuming a chance level of 50%, binomial tests indicated that the kindergarteners selected the complex containing the positive cue of Problem 1 at a level significantly above chance ($p < .01$), while the second graders selected at a level not significantly below chance ($p > .20$).

DISCUSSION

The major result of this experiment was that kindergarten children learned faster under the Same condition than under the Reversal condition, while second-grade children learned at similar rates under the two conditions. This finding suggests that kindergarten children tend to extinguish instrumental choice responses more slowly than mediating responses and that the reverse probably is true of second-grade children. Additional support for this inference is provided by the data indicating that kindergarteners do and second graders do not show a tendency to terminate the run of perseverative responses in Problem 3 with the selection of the complex containing the positive cue of Problem 1. The hypothesis of a developmental change in the relative rates at which mediating and choice responses are acquired and extinguished thus is supported.

The hypothesis can be invoked to explain the increase with age in the proportion of *S*s making optional reversal shifts. As Campione (1970) has pointed out, *S* begins the optional shift problem from a mediational viewpoint with a strong mediating response to the previously relevant dimension and a strong choice response to the cue on that dimension which now is a part of the negative stimulus complex. Since *S* will respond incorrectly on the first few trials, both the prepotent mediating and choice responses will undergo extinction. If the choice response extinguishes faster than the mediating response, *S* will be likely to learn an optional reversal shift. If the mediating response extinguishes faster, then an optional ED shift or inconsistent responding becomes the more likely result.

The Kendler et al. (1962) results are not completely consistent with those of the present study. There is no indication in the present results that even some second-grade children extinguish choice responses slower than mediating responses. Yet Kendler et al. found only approximately 60% optional reversers at a comparable age level. It seems likely, however, that other factors, such as dimensional salience, are involved in the determination of the type of shift which is made. Even under conditions where choice response transfer is eliminated, transfer of learned mediating responses is not the sole determinant. Campione (1970), for example, found that under such conditions over one-third of his second-grade sample made optional ED shifts.

A finding related to the present hypothesis was that the run of perseverative responses to the positive cue of the preceding problem was longer for kindergarten than for second-grade children. This result suggests, as might be expected, that the rate difference is one in which younger children acquire and extinguish instrumental choice responses slower than older children. However, the equivalent learning rates for the two age groups in Problems 1 and 2 and in the Same condition of Problem 3 are not consonant with this interpretation. While it could be proposed that the slower acquisition and extinction of choice responses in kindergarteners is counterbalanced by a faster acquisition and extinction of mediating responses, such a proposal certainly is not intuitively appealing. Perhaps it is best merely to note at this time it is frequently found that younger children learn slower than older children (e.g., Tighe & Tighe, 1966).

Verbal behavior may play a role in the type of developmental change indicated by the present results. As the child becomes older, he may begin to use implicit verbalizations in problem solving (e.g., see Flavell, 1970; Luria, 1961). The implicit verbalizations could allow him to rehearse the associations between specific cues and rewards. The rehearsal could in turn facilitate the elicitation and inhibition of responses to specific cues and, hence, produce the faster acquisition and extinction of instrumental choice responses.

REFERENCES

CAMPIONE, J. C. Optional intradimensional and extradimensional shifts in children as a function of age. *Journal of Experimental Psychology*, 1970 **84**, 296–300.

CARON, A. J. Discrimination shifts in three-year-olds as a function of shift procedure. *Developmental Psychology*. 1970, **3**, 236–241.

DICKERSON, D. J., WAGNER, J. F., & CAMPIONE, J. C. Discrimination shift performance of kindergarten children as a function of variation of the irrelevant shift dimension. *Developmental Psychology*, 1970, **3**, 229–235.

Flavell, J. H. Developmental studies of mediated memory. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in child development and behavior.* Vol. 5. New York: Academic Press, 1970.

Goodwin, W. R., & Lawrence, D. H. The functional independence of two discrimination habits associated with a constant stimulus situation. *Journal of Experimental Psychology*, 1955, **48**, 437–443.

Isaacs, I. D., & Duncan, C. P. Reversal and nonreversal shifts within and between dimensions in concept formation. *Journal of Experimental Psychology*, 1962, **64**, 580–585.

Kendler, H. H., & D'Amato, M. F. A comparison of reversal shifts and nonreversal shifts in human concept formation. *Journal of Experimental Psychology*, 1955, **48**, 165–174.

Kendler, H. H., & Kendler, T. S. Vertical and horizontal processes in problem solving. *Psychological Review*, 1962, **69**, 1–16.

Kendler, H. H., & Kendler, T. S. Mediation and conceptual behavior. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 2. New York: Academic Press, 1968.

Kendler, T. S., Kendler, H. H., & Learnard, B. Mediated responses to size and brightness as a function of age. *American Journal of Psychology*, 1962, **75**, 571–586.

Lawrence, D. H., & Mason, W. A. Systematic behavior during discrimination reversal and change of dimensions. *Journal of Comparative and Physiological Psychology*, 1955, **48**, 1–7.

Luria, A. R. *The role of speech in the regulation of normal and abnormal behavior.* New York: Liveright, 1961.

Spence, K. W. The nature of discrimination learning in animals. *Psychological Review*, 1936, **43**, 427–449.

Tighe, T. J., & Tighe, L. S. Overtraining and optional shift behavior in rats and children. *Journal of Comparative and Physiological Psychology*, 1966, **62**, 49–54.

(Received November 26, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 123–134

FRUSTRATION PHENOMENA IN PAIRED-ASSOCIATE LEARNING[1]

R. A. CHAMPION,[2] T. E. McCANN, AND J. A. RUFFELS

University of Sydney, Australia

Two experiments were designed to detect frustration phenomena hitherto only observed in subhuman animals, frustration being defined as the omission of the response item after a correct anticipation, in a partial reinforcement schedule. Evidence was obtained for (*a*) an improvement in the performance of a dominant response immediately following frustration, (*b*) a general improvement in the performance of other responses during partial reinforcement, and (*c*) the development of activating properties in an initially neutral stimulus (light) consistently paired with frustration.

If the response item in paired-associate (PA) learning acts as a conventional reinforcer, then its omission in a partial reinforcement schedule should qualify as a case of frustration and generate three effects hitherto observed with subhuman animals, viz., (*a*) an improvement in the immediately following performance of a dominant response, which has become known as the frustration effect (FE) (Amsel & Roussel, 1952), (*b*) a more general and slower improvement in overall performance (e.g., Goodrich, 1959), and (*c*) the development of activating properties in an initially neutral stimulus consistently paired with the frustration (Wagner, 1963). The following experiments were designed to test these predictions.

The expectation of improved performance with partial reinforcement in PA learning may seem to be incompatible with the results of at least two studies in which percentage frequency of occurrence of the response member (ORM) of the PAs was an independent variable; using measures such as frequency of correct response and trials to criterion, Goss, Morgan, and Golin (1959) and Schulz and Runquist (1960) found that acquisition performance increased with ORM, whereas the present formulation predicts a decrease. However, Schulz and Runquist observed that reduced ORM increased the frequency with which *S* gave a correct response once the first correct anticipation had been made. The resolution of the apparent discrepancy may therefore lie in the assumption that the response member only acts as a reinforcer when it follows a correct response, and that its function prior to correct anticipation is to provide information (Estes, 1969) or promote stimulus–response contiguity (Champion & Markham, 1967). If this assumption is valid, then, within limits, decreased ORM should cause performance to be weaker early and stronger late in acquisition.

The theoretical background of the present studies therefore features the hypothesis that a reinforcer is only effective if there is something to be reinforced, empirically represented in PA learning when the response follows the stimulus in a correct anticipation. The reinforcement process is thought to be observed when the strength of the response continues to increase beyond the first correct anticipation, so that any event having this effect is defined as a reinforcer. Since the ORM is observed to have such an effect, it is taken to be a reinforcer. Hypotheses about the inherent properties of reinforcers are not deemed relevant to the immediate purpose; in particular, it is not implied that ORM confirms *S*'s reaction or has a pleasurable effect like food may have for a hungry animal. Nor is it supposed that ORM after an incorrect anticipation is punishing, although it should follow from the basic hypothesis that it is at least

---

[1] This research was supported by Grant A65/15488 from the Australian Research Grants Committee to the first author. The third author completed Exp. I as a fourth-year undergraduate honors project, under the supervision of the first author. The second author served as a research assistant for Exp. II.

[2] Requests for reprints should be sent to R. A. Champion, Department of Psychology, University of Sydney, N.S.W. 2006, Australia.

speeds on Probe Presentations 1–5 are shown in Fig. 4. In view of the fact that contrary to prediction, the performance of Group F was not consistently better than that of Group P over later probes, an analysis of variance was conducted in the mean response speeds over Probes 2–5, $F\ (2,57) = 4.91$, $p < .05$; both Groups F and P showed significantly superior performance to Group C, $t\ (57) = 2.91$, $p < .01$, and 2.48, $p < .05$, respectively, but Group F was not significantly superior to Group P, $t\ (57) = .42$, $p > .05$. Although not of statistical significance, Group P was consistently superior to Group F on measures where they were expected to be comparable (Fig. 2 and 3). Since any such difference could mask the predicted superiority of Group F in this third case, a difference score was calculated for each *S* by subtracting the mean score for the general effect (Fig. 3) from the mean score for Probes 2–5 (Fig. 4). In these terms, Group F was superior to combined Groups P and C, $t\ (58) = 2.33$, $p < .05$.

EXPERIMENT II

Experiment I provided clear evidence for an FE in PA learning, but the more

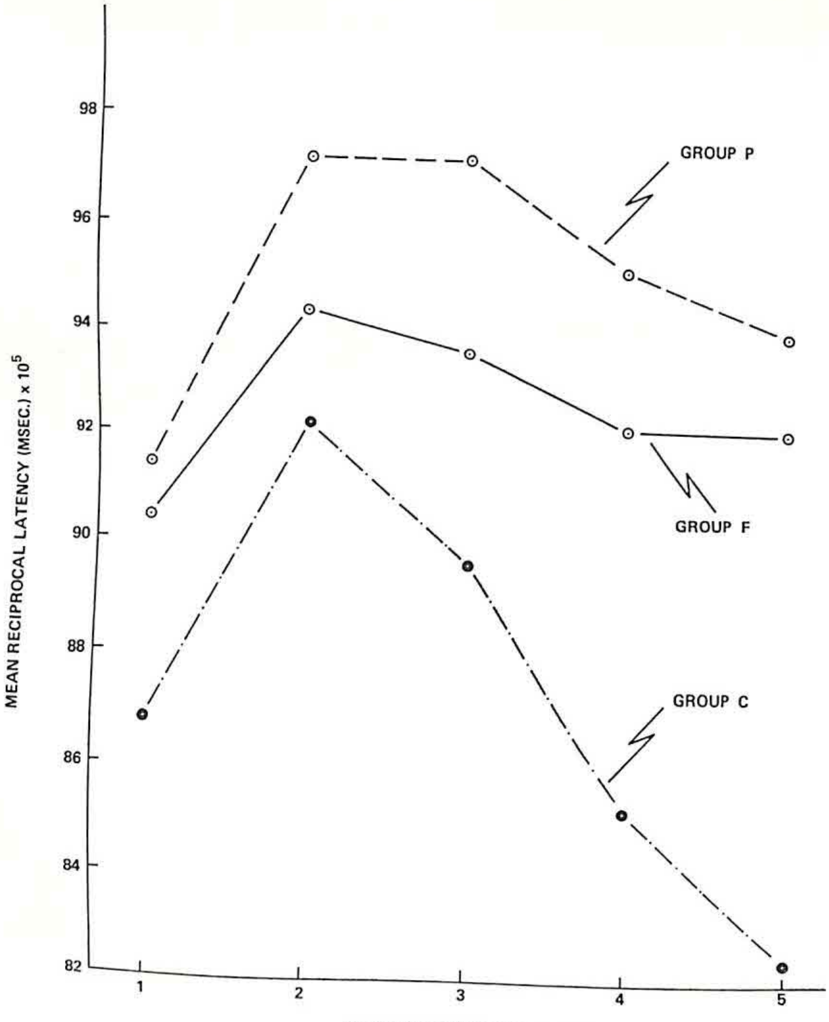


FIG. 2. Mean response speeds on presentations immediately following non reinforced responses in Groups F and P and on corresponding reinforced responses in Group C, Exp. I.

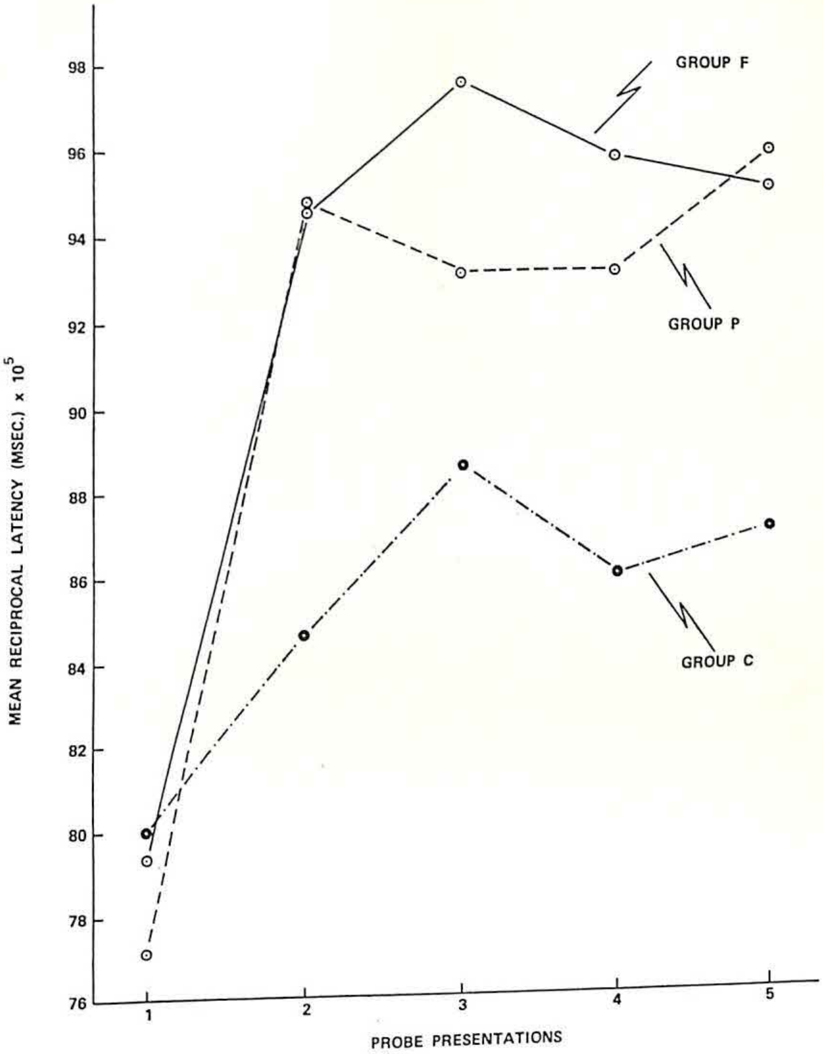


Fig. 3. Mean response speeds on presentations immediately following reinforced responses, Exp. I.

general effect merely approached statistical significance and the status of any acquired activational effect was equivocal.

Method

In an attempt to make the second and third effects occur more strongly, the method of the first study was modified in five main ways in Exp. II: (*a*) A partial reinforcement schedule was imposed on the experimental groups from the outset of training, instead of all groups having an initial period of continuous reinforcement. (*b*) More difficult PAs were used, in the form of letters and nonsense syllables in place of commonly associated words. (*c*) Following the procedure of Wagner (1963), the initially neutral stimulus (the light) was flashed between presentations in Group P rather than forming part of a randomly determined presentation. (*d*) On nonreinforced presentations in Group F, the light was flashed immediately before the response item would have appeared if it had been a reinforced presentation (S-L presentations in Fig. 5), instead of occurring immediately before the stimulus item. (*e*) The probe response was simplified by having *S* call out the stimulus item as it appeared, rather than having to give a response learned earlier in training.

Subjects.—There were 68 *S*s, of the same kind as those used in Exp. I.

Apparatus.—With the following exceptions the apparatus was identical with that used in Exp. I. The *S* was seated in a sound-reduced room separate from that containing *E* and the main equipment. The microphone used by *S* for intercommunication also activated a voice key to stop the chronoscope. The PAs were: J-ZOK, F-XUP, G-VEF, B-DAJ, D-YIS, H-GEX.

Procedure.—The *S*s were allocated to four groups at random within quadruplets, giving 17 in each group. In Stage 1 of the procedure, all *S*s were given 24 trials. Each trial involved the presentation of each of the six stimulus items, with or without the response item, along with three "blank" presentations consisting of the light alone or no stimulus at all, according to the group conditions; these nine presentations occurred in a predetermined random order that varied from trial to trial. Other than on blank presentations, the stimulus word alone appeared for 3 sec. (S presentations in Fig. 5) or else it was followed by the response word for .5 sec., with an interval of 3.5 sec. between the onsets of the two items. The interpresentation interval remained at 10 sec. In Group F, the response item was omitted on half of the presentations, when the light was flashed for .5 sec. immediately following the offset of the stimulus item (S-L presentations in Fig. 5). Group P was on the same reinforcement schedule as Group F; the light also occurred just as often, but it was flashed on blank presentations (L presentations in Fig. 5). In Groups CF and CP, the response item always followed the stimulus item, except on probe presentations, but the light schedules corresponded to those for Groups F and P, respectively.

Stage 2 of the procedure consisted of a further six trials in which the four groups were all run under the same conditions. The response item occurred on every occasion (continuous reinforcement) and there were no blank presentations.

On the last presentation of Trials 3, 5, 9, 11, 13, 15, 19, 21, 23, 25, 27, and 29, the stimulus item was used as a probe, being preceded by the light for 3.5

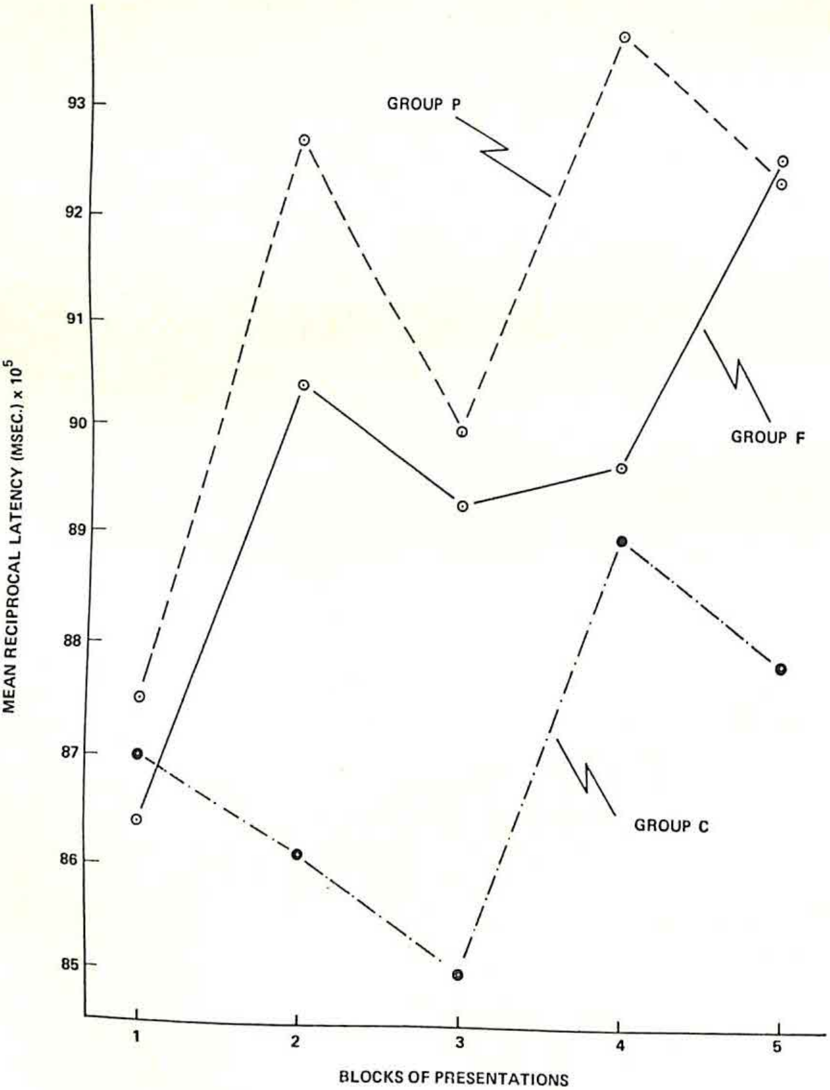
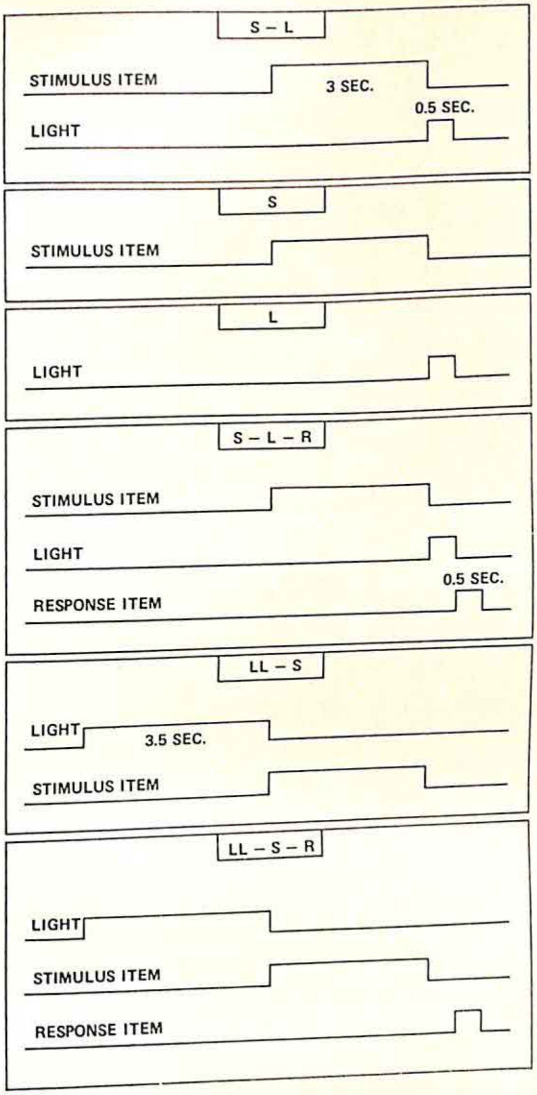


FIG. 4. Mean response speeds on probe presentations, Exp. I.



GROUP	TYPE	APPROX. %
STAGE 1		
F	S – R	32
	S–L	32
	BLANK	32
P	S – R	32
	S	32
	L	32
CF	S – R	32
	S – L – R	32
	BLANK	32
CP	S – R	64
	L	32
PROBES	LL – S	4
STAGE 2		
ALL	S – R	91 2/3
PROBES	LL–S–R	8 1/3

FIG. 5. Presentation types and reinforcement schedules in Exp. II (S–R presentation types are not shown).

sec. (LL-S and LL-S-R presentations in Fig. 5). Each stimulus item was used as a probe twice. On Trials 3–23 (Stage 1), the response item was omitted on probe presentations for all *S*s. Within the limits already implied, there was the general restriction that no more than two consecutive presentations be identical.

As indicated in the following instructions, the session began with two practice trials to ensure that *S* followed directions. Since the chronoscope stopped when *S* read out the stimulus word, these response latencies only allowed a general performance measure, leaving frequency of correctly anticipated following nonsense syllables as the only available measure of PA learning.

Instructions.—The instructions were as follows:

> This experiment is concerned with the learning of nonsense syllables as responses to letters. There are six letter–nonsense-syllable pairs and each pair will be presented a number of times. In each pair the letter will be presented on the left-hand screen. Your task is to learn the nonsense syllable that is paired with each letter. The nonsense syllable paired with the letter will appear on the right-hand screen. As soon as you see the letter on the left-hand screen, read it out as quickly as possible and then say the nonsense syllable that you think is associated with it. Sometimes (*Groups F and P only*: the nonsense syllable paired with the letter will not appear. Also) the time between presentations of each pair may vary, that is, the time between one trial and the next. Therefore you will have to keep your eyes on the screen so that you will be ready to respond as soon as the letter is presented. The light will

come on at various times during the experiment. Remember that you have to call out the letter as soon as it appears and then give the associated nonsense syllable as quickly as possible. Have you any questions so far?

We will now have a couple of practice trials to introduce you to the items you will have to learn and also to the format of the experiment. With these trials, I want you to read out the letter and then read out the nonsense syllable when it appears. And read it out as a syllable, please, not letter by letter. (Two trials, with stimulus and response items on each presentation). Okay, do you get the idea? Remember that during the experiment (*Groups F and P only*: the nonsense syllable will not always appear. Still), as soon as the letter appears you are to read it out and say the associated nonsense syllable as quickly as possible. Do not wait for the nonsense syllable to appear before calling it out. At first you may simply be guessing, but the task is easy and you will learn it quickly. Remember to respond loudly, quickly, and clearly each time.

Results

Mean numbers of correct anticipations given by Groups F and P (partial reinforcement) and CF and CP (continuous reinforcement) are represented in Fig. 6. Following Schulz and Runquist (1960), mean Correct/Opportunity ratios were calculated for each condition by expressing the number of correct anticipations following the first reinforced correct response as a proportion of the number of following opportunities to give a response. The obtained values were .94 and .88 for combined Groups F and P and combined Groups CF and CP, respectively, but the difference was not statistically significant, $t\,(66) = .95, p > .05$.

Frustration effect.—During Trials 14–25, when *S*s were making relatively few errors, there were 27 presentations (excluding probes) immediately following a reinforced presentation in Groups F and P, and 27 corresponding nonreinforced presentations in Groups CF and CP. The data from these presentations are shown in Fig. 7. The mean performance of Groups F and P

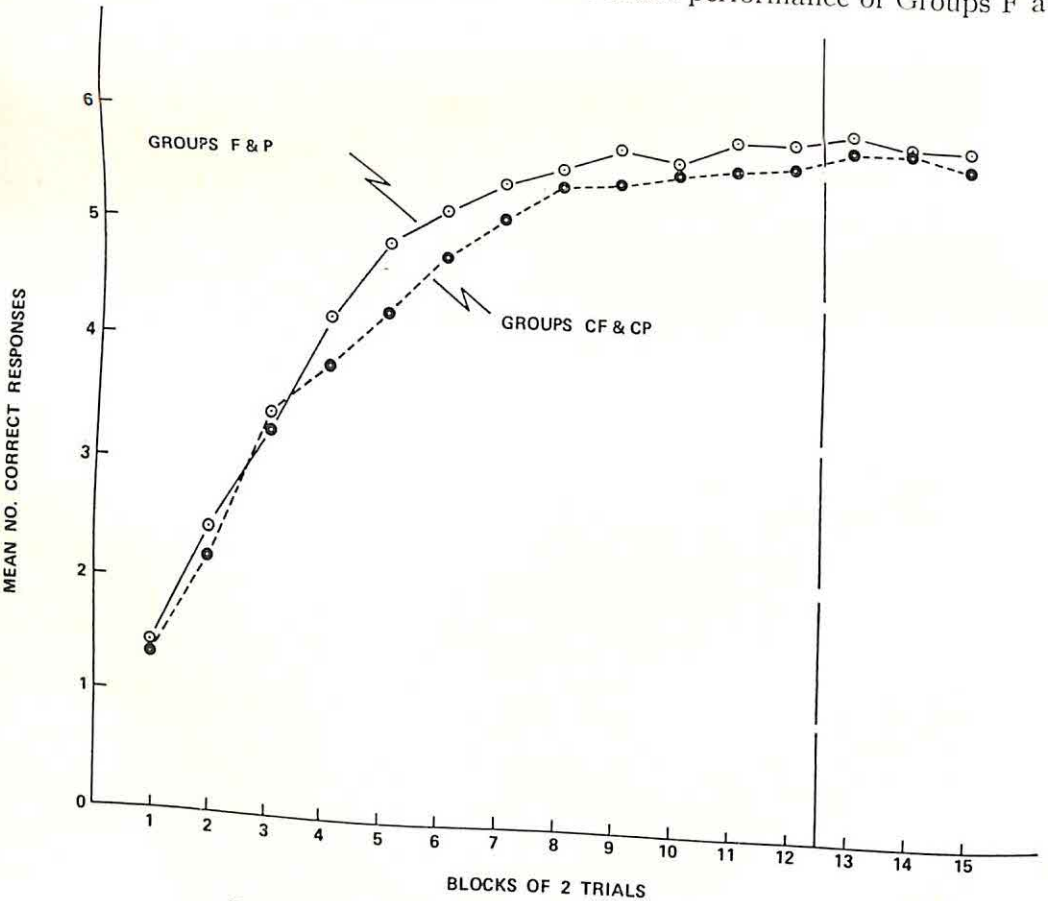


FIG. 6. Mean frequency of correct response, Exp. II.

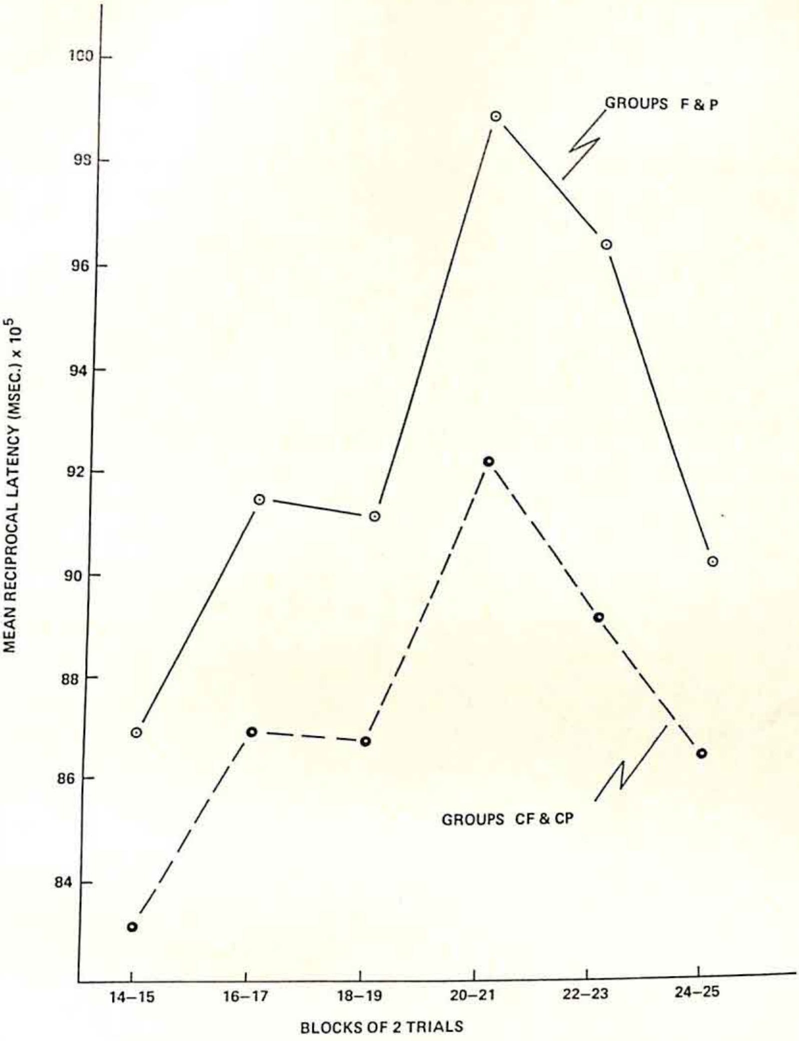


FIG. 7. Mean response speeds on presentations immediately following nonreinforced responses in Groups F and P and on corresponding reinforced responses in Groups CF and CP, Exp. II.

was significantly superior to that of Groups CF and CP, F (1, 64) = 5.41, $p < .05$.

General effect.—Over the same set of trials there were 35 presentations (excluding probes) when S was virtually certain to have given a correct, reinforced response on the immediately preceding presentation. The mean speed data for these presentations, excluding the few occasions on which the immediately preceding response was incorrect, are shown in Fig. 8. The performance of Groups F and P was significantly superior to that of Groups CF and CP, F (1, 64) = 4.44, $p < .05$. This significant difference did not persist in Stage 2, when all *S*s were treated alike, F (1, 64) = 3.49, $p > .05$.

Acquired activation.—There was no general effect (Groups F and P vs. CF and CP) on probe presentations at the end of Stage 1 (Probes 8–9) or Stage 2 (Probes 11–12), F (1, 64) = 1.58 and 1.52, respectively, $p > .05$. This finding allowed a comparison of Group F with combined Groups P, CF, and CP in the knowledge that any observed difference in favor of Group F could be interpreted as due to acquired activation instead of a general activational effect. Such a difference was found significant on Probes 8–9, t (64) = 2.05, $p < .05$, but not Probes 11–12, t (64) = 1.67, $p > .05$ (Fig. 9). Tests of linear trends over Probes 1–9 in Group F and Groups P, CF, and CP (separately or combined) proved significant only for Group F, F (1, 144) = 10.10, $p < .01$.

Discussion

The aims of Exp. II were met, in that it not only replicated the FE observed in Exp. I but also provided relatively unequivocal evidence for a general activational effect due to frustration and for an acquired activational effect. The test of the general activational effect was a rigorous one, confined as it was to presentations immediately following a correct response, but the partially reinforced groups still showed significantly superior performance. The existence of a significant trend over probe presentations only in Group F, at the exclusion of the three other groups, is strongly suggestive of an acquired activational effect, despite the loss of a significant difference in Stage 2.

CONCLUSION

It is likely that the FE and the general activational effect are two cases of the same phenomenon. The FE is only observed after the correct response is the rule, when frustration is a meaningful operation by way of the omission of reinforcement. As seen in learning by subhuman animals (e.g., Goodrich, 1959), the general effect appears to develop slowly because it also probably awaits the prior development of the correct response. The fact that activational effects due to frustration can become attached to neutral stimuli present at the time, in human as well as subhuman Ss, must have implications for the explanation of everyday human behavior. Most important,

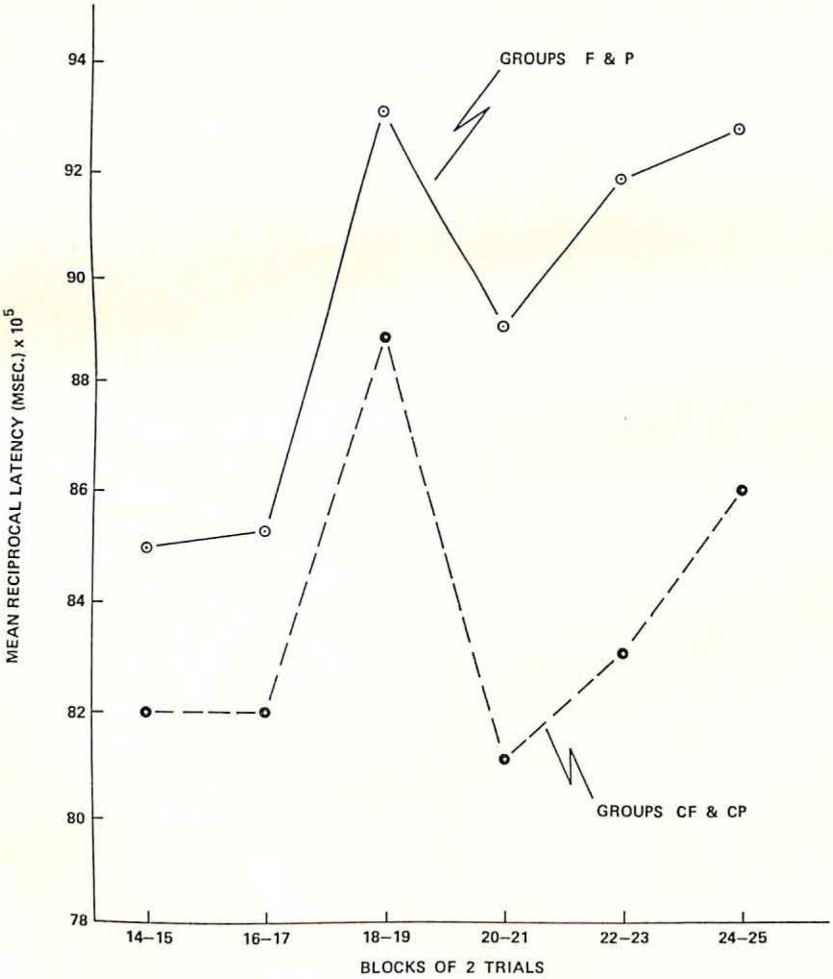


FIG. 8. Mean response speeds on presentations immediately following reinforced responses, Exp. II.

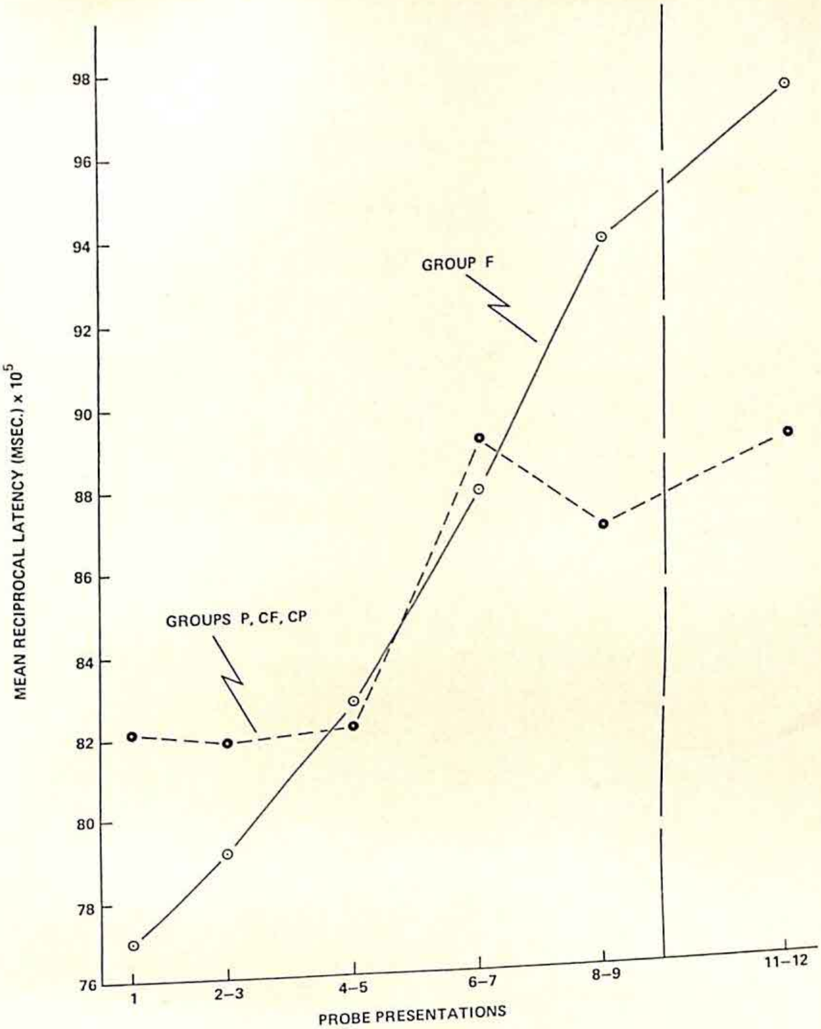


FIG. 9. Mean response speeds on probe presentations, Exp. II.

Perhaps, the findings of the present experiments contradict the assumption that data from lower animals have no direct relevance to human activity (e.g., Kintsch, 1970, p. 40), being more in keeping with the results of workers like Nelson (1971).

The present data also confirm the deduction that the response item in PA learning acts as a reinforcer when it occurs after a correct response. While this finding does nothing to settle the issue as to whether the reinforcer promotes the associative process as such or merely has an activational effect, it does seem that its omission raises the organism's level of activation, a reaction that can itself be learned. It must be emphasized that the significant data of Exp. I and II were obtained when correct responses predominated, so that observed effects amounted to increments in performance. It may equally follow that the imposition of frustration in complex situations will yield decrements in performance (e.g., Schmeck, 1970; Schmeck & Bruning, 1968).

REFERENCES

AMSEL, A., & ROUSSEL, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *Journal of Experimental Psychology*, 1952, **43**, 363–368.

CHAMPION, R. A., & MCBRIDE, D. A. Activity during delay of reinforcement in human learning. *Journal of Experimental Psychology*, 1962, **63**, 589–592.

CHAMPION, R. A., & MARKHAM, R. H. Reinforcement and S-R contiguity in paired-associate learning. *Australian Journal of Psychology*, 1967, **19**, 209–216.

Estes, W. K. New perspectives on some old issues in association theory. In N. J. Mackintosh & W. K. Honig (Eds.), *Fundamental issues in associative learning*. Halifax, Nova Scotia: Dalhousie University Press, 1969.

Goodrich, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *Journal of Experimental Psychology*, 1959, **57**, 57–63.

Goss, A. E., Morgan, C. H., & Golin, S. J. Paired-associates learning as a function of percentage of occurrence of response members (reinforcement). *Journal of Experimental Psychology*, 1959, **57**, 96–104.

Kintsch, W. *Learning, memory, and conceptual processes*. New York: Wiley, 1970.

Nelson, T. O. Extinction, delay and partial-reinforcement effects in paired-associate learning. *Cognitive Psychology*, 1971, **2**, 212–228.

Schmeck, R. R. Error-produced frustration as a factor influencing the probability of occurrence of further errors. *Journal of Experimental Psychology*, 1970, **86**, 153–156.

Schmeck, R. R., & Buring, J. L. Task difficulty and the frustration effect. *Journal of Experimental Psychology*, 1968, **78**, 516–520.

Schulz, R. W., & Runquist, W. N. Learning and retention of paired adjectives as a function of percentage occurrence of response members. *Journal of Experimental Psychology*, 1960, **59**, 409–413.

Wagner, A. R. Conditioned frustration as a learned drive. *Journal of Experimental Psychology*, 1963, **64**, 142–148.

(Received November 29, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 135–139

INTERACTION OF AROUSAL AND NUMBER OF LEARNING TRIALS IN PAIRED-ASSOCIATE LEARNING

JOHN W. OSBORNE[1]

University of Alberta

The interaction of arousal and number of learning trials in paired-associate recall was investigated in two experiments. The initial purpose was to test the validity of the repetition bias. The facilitation of high-arousal recall by multiple massed learning trials in Exp. I led to the investigation, in Exp. II, of whether this effect was primarily a function of repetition of the learning material or the time taken by more than one learning trial. Results of Exp. II suggested that the passage of time was the more critical influence upon recall. The findings of both experiments were interpreted as being equivocal in regard to neural consolidation theory.

Although it is generally believed that repetition of a list during learning facilitates its recall, (cf. Tulving & Madigan, 1970), the "action decrement" theory of Walker (1958) maintains that while consolidation of the memory trace is occurring there is a negative bias against repetition. Repetition of the material to be learned produces higher levels of neural reverberation which, although leading to greater ultimate memory, inhibits short-term memory. This situation is accentuated if the material is learned under high-arousal conditions. Several paired-associate (PA) learning studies have confirmed the consolidation hypothesis by showing that low-arousal recall is greater than high-arousal recall over the short term while the reverse occurs over the long term (Kleinsmith & Kaplan, 1963, 1964; Walker & Tarte, 1963).

As an alternative to electroconvulsive shock (ECS) techniques, Walker (1967) has suggested the presence of the repetition bias in verbal learning studies as an index of the length of the consolidation period. It is presumed that when repetitive interference ceases the consolidation period has ended. Available evidence is equivocal as far as this problem is concerned, with ECS studies suggesting that the consolidation period lasts from 5 sec. to several days while verbal learning research suggests 20 to 45 min. (cf. Walker, 1967).

In exploring this problem further, Tarte (1964) varied the number of learning trials for high- and low-arousal PAs as well as the spacing between trials prior to a test of recall 1 wk. later. The intention was to determine the effect of either a second learning trial or an interference trial, following the first learning trial, upon performance by introducing it at different stages of the consolidation period and thus shedding some light on the duration of that period. The interference trial was composed of the same words used in the first trial but paired with different numbers. Tarte expected repetition to interfere decreasingly with recall performance, reaching a plateau between 20 and 45 min. Inhibition of recall following an interference trial was expected to increase to a plateau between the same intertrial intervals.

Results of this study confused rather than clarified the problem under investigation. Unpredicted increases and decreases in recall for all conditions occurred. The repetition bias did not materialize in the way predicted from earlier studies.

High-arousal recall after two learning trials decreased as the spacing between trials increased. Performance decreased to a plateau between the intervals of 6 min. and 45 min. Contrary to prediction, high-arousal recall after one learning trial plus a spaced interference trial did not decrease to a plateau between the intertrial intervals of 20 min and 45 min. High-arousal recall actually decreased between these intervals. The results were generally quite unlawful as far as consolidation theory is concerned.

Tarte's (1964) study employed a retention interval of 1 wk. in all conditions so that the validity of the repetition bias was not completely tested. The question remains as to whether data will support the implication of a positive linear relationship between the number of learning trials and the "degree of temporary inhibition [p. 113]" referred to by Walker (1958) as resulting from the repetition bias in short-term recall of high-arousal material. The idea of repetition failing to increase recall performance contradicts most available data; however, the action decrement theory remains to be adequately tested in this regard. If, in fact, the relationship between inhibition of recall and number of learning trials is not linear for high-arousal material, other questions arise such as how many learning trials produce optimal short-term recall and what factors are responsible for any departures from the action decrement theory.

---

[1] Requests for reprints should be sent to John W. Osborne, Department of Educational Psychology, University of Alberta, Edmonton, 7, Alberta, Canada.

Previous research has shown that arousal inhibits recall when it is tested 2 min. after one learning trial begins; however, the effect of multiple massed trials has not been studied. According to Walker's, action decrement theory, massed learning trials should increase perseveration and reverberation of the memory trace, thus inhibiting immediate recall. The purpose of this study is to investigate this hypothesis by examining the interaction of arousal and number of learning trials on a test of short-term recall of low- and high-arousal PAs.

EXPERIMENT I

Method

Subjects.—The *S*s were 48 unpaid volunteers from educational psychology courses at the University of Houston.

Materials.—The learning materials consisted of two lists of PAs (cf. Walker & Tarte, 1963). These were eight high-arousal words (KISS, RAPE, VOMIT, SLUT, EMBRACE, MONEY, PASSION, SEX) and eight low-arousal words (PENCIL, CARROT, BERRY, WALK, FLOWER, POND, WHITE, GLASS). Single digits (ranging from 2 to 9) were randomly assigned as the response terms in each list. Color slides were employed to separate arousal effects between presentations of the PAs (cf. Kleinsmith & Kaplan, 1963).

Design.—The study used a 2 × 3 independent groups factorial design. The factors were (*a*) a high- and low-arousal list of eight PAs each and (*b*) one, two, or three learning trials. The dependent variable was the number of PAs correctly recalled on an immediate test of retention. The *S*s were randomly assigned to one of the six experimental groups. The eight *S*s in each group learned a different order of presentation of the words for both lists. To control for serial position effects, each word appeared once in each position in each list. The *S*s saw the same list order for one, two, or three trials during learning, while all *S*s saw a different list order for recall. Recall lists were varied in the same way as learning lists to control for serial position.

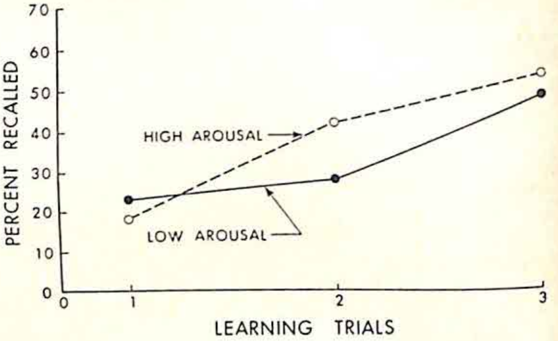


FIG. 1. Mean recall scores for low- and high-arousal PAs over learning trials.

Procedure.—The learning and recall procedure was similar to that used by Kleinsmith and Kaplan (1963). The *S*s were given either one, two, or three continuous learning trials with high- or low-arousal lists which were presented by means of a slide projector with an exposure time of 5 sec. Individual *S*s first saw the word alone, then paired with a single digit response. Interpolated color naming (two slides) was used to separate the arousal effects of each pair in the list. The *S* was told to "concentrate carefully on calling out the word number pairs and naming the colors" but was not specifically told that he would be tested for recall.

On a test of immediate recall, *S* was asked to recall the missing number for each stimulus word as it was presented. An exposure time of 5 sec. and interpolated color naming were used as before.

Results

Figure 1 shows the mean recall scores for the low- and high-arousal PAs over the trial conditions. Although the recall of low-arousal PAs after one trial was greater than the recall of high-arousal PAs, as consolidation theory predicts, the difference was not significant. The situation was reversed after two and three trials, showing that if greater perseveration of the memory trace resulted from additional trials it did not increasingly inhibit short-term recall. Contrary to the repetition bias, high-arousal recall scores increased with trials like most recall data recently reviewed (Tulving & Madigan, 1970).

An analysis of variance of the recall scores for the six experimental conditions shown in Fig. 1 revealed a significant main effect due to trials, $F\ (2,\ 42) = 49.0$, $p < .001$. A nearly significant interaction, $F\ (2,\ 42) = 3.1$, $p < .06$, showed that arousal and number of learning trials interacted to affect retention. The greater effect of arousal after two learning trials observed in Fig. 1 was confirmed as statistically significant by a post hoc comparison using the Tukey procedure, $q\ (6,\ 42) = 4.2$, $p < .05$.

Examination of the serial position data for low- and high-arousal PA recall after one, two, and three trials relative to item position at the time of learning and recall revealed no systematic effects in relation to the manipulation of arousal or the number of trials. There were only slight and infrequent departures from the usual symmetric serial position effects.

Discussion

The results of this study indicate that arousal inhibits recall after one learning trial but facilitates recall after two and three trials. Both low- and high-arousal lists showed an increase in recall as trials increased. The repetition bias for high-arousal material did not appear.

The unmentioned, but critical, variable so far for consolidation theory is time. Obviously it is confounded with other variables involved in repetition of the learning lists. In discussing the action decrement theory, Walker (1958) talks of a "degree of temporary inhibition" caused by the repetition bias. The present results suggest that either this bias was absent or that the degree of temporary inhibition is small or of brief duration.

According to consolidation theory (Hebb, 1958), repetition of the list of PAs increases perseveration and reverberation of the memory trace while at the same time providing additional time for consolidation. The results of this experiment suggest that optimal high-arousal recall may be a function of not only trials but time taken by the trials. Perhaps the repetition bias is inversely related to the number of trials rather than positively. This explanation would make the observed interaction between arousal and trials more understandable.

TABLE 1

HYPOTHESES AND PLANNED COMPARISONS FOR MEAN PA RECALL

No.	Hypothesis	$t(14)$
1	HA, 1T+ > HA, 1T	1.79*
2	HA, 2T > HA, 1T+	.91
3	LA, 1T > LA, 1T+	1.79*
4	LA, 2T > LA, 1T+	1.45
5	LA, 1T > HA, 1T	2.89*
6	HA, 1T+ > LA, 1T+	1.41
7	HA, 2T > LA, 2T	1.38
8	HA, 2T > HA, 1T	2.90*
9	LA, 2T > LA, 1T	−.37
10	HA, 2T = HA, 1T+	.91

Note.—HA = high arousal, LA = low arousal, 1T = one trial, 2T = two trials, + = plus interpolated color naming.
* $p < .05$ (one-tailed test).

The validity of this interpretation cannot be tested in the present study because the effects of passage of time and repetition in the learning trials are confounded. Consequently, a second study attempted to solve this problem.

EXPERIMENT II

This study attempted to isolate the effects of time and repetition of the stimulus within the context of multiple learning trials in an attempt to determine whether the additional time for consolidation of the memory trace after one learning trial or the repetition of the stimulus over two learning trials was the most significant component of the interaction of arousal and learning trials found in Exp. I. Because arousal had its greatest effect after two trials in Exp. I, it was decided to use one and two trials in Exp. II.

The hypotheses to be tested in Exp. II are listed in Table 1.

If according to consolidation theory the passage of time rather than repetition of the learning material interacts with arousal to facilitate recall, then longer learning periods should favor high-arousal recall over low-arousal recall and vice versa for immediate recall (Hypotheses 1, 5, 6, 7, 10).

If repetition of the list is more critical than the time taken in learning, then increased trials can be expected to produce an increase in recall for both high- and low-arousal lists (Hypotheses 2, 4, 8, 9).

If, in contrast to high-arousal material, low-arousal material is subject to normal

TABLE 2

MEAN PA RECALL IN PERCENTAGES FOR AROUSAL × LEARNING TRIAL CONDITIONS

1 Trial		1 Trial +		2 Trials	
LA	HA	LA	HA	LA	HA
23.4	14	14	23.4	21.8	29.6

Note.—LA = low arousal, HA = high arousal, + = interpolated color naming.

forgetting over time, then interpolated activity between learning and recall should inhibit recall performance (Hypotheses 3, 4).

Method

Subjects.—The *S*s were 48 unpaid volunteers from educational psychology courses at the University of Houston.

Materials.—Materials were the same as those used in Exp. I except that additional color slides were used for the interpolated activity of color naming.

Design.—A modified 2 × 2 independent groups factorial design was used. The factors were (*a*) low- and high-arousal lists of PAs and (*b*) one or two learning trials. Two control groups were also used. One control group learned the high-arousal list for one trial, then named color slides for 2½ min. while the other group did the same for the low-arousal list. The assignment of *S*s to groups and the presentation of lists were the same as Exp. I.

Procedure.—The procedure was the same as in Exp. I, except that the two control groups named colors after one learning trial until the same amount of time had elapsed as was taken for two learning trials (5 min.).

Results

An analysis of variance for the effect of arousal and trials upon PA recall produced a significant main effect for trials, F (1, 28) = 4.6, $p < .05$, and a significant interaction between arousal and trials, F (1, 28) = 7.0, $p < .05$. This indicates the possibility of the time taken by two trials as sufficient to produce the change in the direction of the arousal effect seen in Exp. I.

Table 2 compares the mean recall scores for low- and high-arousal lists of PAs after one trial, one trial plus color naming, and two trials, while Table 1 shows the results of the planned comparisons expressed in earlier hypotheses.

As predicted, low-arousal recall was greater high-arousal recall at one trial, while the reverse tended to be true at one trial plus color naming. Holding the arousal conditions constant and increasing the retention interval produced the characteristic interaction reported by Kleinsmith and Kaplan (1963). The additional time taken for color naming may have produced forgetting of the low-arousal list and consolidation of the memory trace for the high-arousal material. After two learning trials recall of the high-arousal material improved significantly, whereas there was slight drop in the recall of low-arousal material. Recall of both low- and high-arousal lists at two trials was superior to their recall at one trial plus color naming; however, the differences were not statistically significant.

Contrary to expectation, low-arousal recall was lower after two trials than after one, although higher than after one trial plus color naming. This suggests that the additional learning trial was unable to offset the retention loss occurring after one trial. Apparently the facilitation of retention from the extra trial was less than the retention loss shown in the one-trial-plus-color-naming condition due to the passage time.

Low-arousal recall after one trial plus color naming was significantly less than after one trial. As predicted, the passage of time facilitated the forgetting of low-arousal material.

High-arousal recall after two trials was significantly greater than after one trial. Whether it was the extra time or repetition of the list provided by the second trial that was responsible for the improved performance can be determined by comparing high-arousal recall after one trial plus color naming and after two trials. The lack of a significant difference suggests, once again, that the passage of time after one trial rather than the repetition of the list was the critical factor. This interpretation is supported by the significant difference between high-arousal recall at one trial and one trial plus color naming.

As in Exp. I, the serial position of items on learning and recall lists did not appear to be associated with systematic changes in arousal or the number of trials.

DISCUSSION

The significantly greater recall of low-arousal material in comparison to high-arousal recall after one trial and the fact that, although not significant, after one trial plus interpolation there was a trend for high-arousal recall to be greater than low-arousal recall support the importance of the passage of time during learning as the crucial factor affecting high-arousal recall.

The fact that high-arousal recall after two trials was significantly greater than after one trial argues against the action decrement theory. According to the theory, an extra trial should increase perseveration of the memory trace resulting in inhibition of short-term memory. Paradoxically, low-arousal recall over the same condition behaved in a way expected of high-arousal recall. The failure of low-arousal recall to increase with trials is difficult to explain unless it was a result of experimental error made more critical by small cell size in each condition.

The increase in high-arousal recall with extra learning trials seems to be a stable result, in that it occurred in both studies. This argues strongly against the repetition bias.

The failure to find a significant difference between high-arousal recall after one trial plus interpolation and two trials suggests that extra time is the most important reason why high arousal after two trials is greater than after one. Perhaps one presentation of the list creates optimal reverberation so that additional presentations of the list do not significantly affect the reverberation process. This suggests a type of overflow or safety-valve mechanism. The significantly greater recall of high-arousal material after one trial plus interpolation than after one trial also points to the importance of time.

The fact that low-arousal recall after one trial was significantly greater than after one trial plus interpolation confirms the hypothesis of low-arousal material being subject to normal forgetting over time.

The failure to find a significant difference between low-arousal recall after one trial plus interpolation and two trials is a problem for the theory of the differential effect of time upon arousal; however, there was a trend in this direction. Once again, the unpredicted behavior of low-arousal recall makes interpretation of high-arousal recall performance more tentative.

Evidence from this study is equivocal as far as consolidation theory is concerned. Results argue strongly against simple operationalization of the action decrement theory. The repetition bias failed to appear. Additional learning trials actually increased high-arousal recall. The findings were compatible with an interpretation of time as the critical variable in the arousal-trials interaction. The failure of the second trial to significantly increase high-arousal recall over high-arousal recall for one trials plus interpolation suggests that the additional learning trial is primarily an extension of the retention interval which permits increased consolidation of the memory trace. To this extent the data support the notion of consolidation over time.

In view of the above findings, time seems to be the most likely explanation for the interactions between arousal and trials. If time is the critical variable for arousal in these two experiments, then the obtained interactions are compatible with those obtained by other *E*s. However, the slight differences between high- and low-arousal recall in Exp. I after one trial and after two trials in Exp. II, as well as the absence of a main arousal effect in both studies, make these interactions suspect inspite of the fact that differences were in the directions predicted by consolidation theory.

REFERENCES

HEBB, D. O., *A textbook of psychology*. Philadelphia: Saunders, 1958.

KLEINSMITH, L. J., & KAPLAN, S. Paired-associate learning as a function of arousal and interpolated interval. *Journal of Experimental Psychology*, 1963, **65**, 190–193.

KLEINSMITH, L. J., & KAPLAN, S. Interaction of arousal and recall interval in nonsense syllable paired-associate learning. *Journal of Experimental Psychology*, 1964, **67**, 124–126.

TARTE, R. D. *The effects of time on reactivation or interference on trace consolidation in verbal learning*. (Doctoral dissertation, University of Michigan). Ann Arbor, Mich.: University Microfilms, 1964, No. 65–5948.

TULVING, E., & MADIGAN, S. A. Memory and verbal learning. *Annual Review of Psychology*, 1970, **21**, 437–484.

WALKER, E. L. The action decrement and its relation to learning. *Psychological Review*, 1958, **65**, 129–142.

WALKER, E. L. Arousal and the memory trace. In D. P. Kimble (Ed.), *The organization of recall*. New York: New York Academy of Sciences, 1967.

WALKER, E. L., & TARTE, R. D. Memory storage as a function of arousal and time with homogeneous and heterogeneous lists. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 113–119.

(Received November 30, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 140–146

IMAGERY AND FREQUENCY PROCESSES IN VERBAL DISCRIMINATION LEARNING[1]

EDWARD J. ROWE[2]

University of Western Ontario

Three experiments are reported in which the relationship between word imagery and situational frequency was investigated in verbal discrimination learning (VDL). In Exp. I, *S*s learned a mixed list of high-imagery (High-*I*) or low-imagery (Low-*I*) pairs under either instructions to repeat the right (R) item of each pair aloud more frequently than the wrong (W) item, or uninstructed conditions. In Exp. II, lists in which the R or the W items were repeated in different pairs were learned under an imagery instruction or control condition. The third experiment compared repeated-right and repeated-wrong list conditions as a function of the *I* value of pairs. The results showed that imagery and frequency factors have independent and additive effects in VDL. The effect of imagery, viewed as lying outside the scope of existing frequency theory, was interpreted in terms of a differential encoding response to the R and W items of a pair.

In verbal discrimination learning (VDL), *S* is typically required to learn which member of each pair of presented verbal items has been designated as "correct" by *E*. Among existing theories of VDL, the most comprehensive is frequency theory (Ekstrand, Wallace, & Underwood, 1966), which accounts for correct discrimination performance on the basis of the differential frequency acquired by right (R) and wrong (W) items of each pair with repeated exposure to the list. Several kinds of "responses," i.e., perceptual, pronunciation, and associative, are assumed to contribute to the frequency of both R and W items during list learning. It is further assumed that *S*s implicitly rehearse the R items, which thereby acquire a higher frequency than the W items, thus producing the cue for discrimination.

There can be little doubt that situational word frequency is a powerful determinant of VDL performance. Studies in which frequency has been experimentally manipulated by repetition of R and W items in different pairs within the list (Ekstrand et al., 1966; Underwood & Freund, 1969), prior free learning or familiarization trials (Lovelace, 1969; Underwood & Freund, 1968), interpair associative similarity (Ekstrand et al., 1966; Fulkerson & Kausler, 1969), and instructions to overtly rehearse the R items (Rowe & Paivio, 1971a; Underwood & Freund, 1968) have almost without exception supported this position.

On the other hand, it has become clear that frequency is but one of a number of word attributes which may affect VDL. This is apparent from the results of several studies, including those in which responding on the basis of a frequency cue has been obviated by the experimental design. For example, Kausler and Boka (1968) examined learning of a double function list, where each R item appeared as a W item elsewhere in the list, and vice versa. This condition markedly depressed performance, but the fact that a double function list could be learned, albeit with difficulty, shows that a discrimination can be made using attributes other than frequency when this is necessitated by the task requirements.

---

[1] This report is based on a dissertation submitted to the Faculty of Graduate Studies, University of Western Ontario, in partial fulfillment of the requirements for the PhD degree. The author is indebted to Allan Paivio for his suggestions and advice as dissertation supervisor. The research was supported by grants to Allan Paivio from the National Research Council of Canada and the University of Western Ontario Research Fund. During the term of the research, the author held a postgraduate scholarship from the National Research Council. The manuscript was prepared while the author held an Izaac Walton Killam Postdoctoral Fellowship at the University of Alberta.

[2] Requests for reprints should be sent to E. J. Rowe, who is now at the Department of Psychology, Memorial University of Newfoundland, St. John's, Newfoundland, Canada.

The nature of many of these attributes which may be used in addition to, or in lieu of, frequency has yet to be specified, but some statements can be made about the relative importance of several variables of this type. Paivio and Rowe (1970) and Rowe and Paivio (1971b) examined the relative effectiveness of rated imagery (*I*), Thorndike-Lorge frequency, and associative meaningfulness in VDL, and found that, of the three, only *I* produced a consistent effect, with High-*I* pairs being easier to learn than Low-*I*. In addition, the results of a postexperimental questionnaire which investigated *S*s' use of four types of learning strategies (Rowe & Paivio, 1971b) suggested that strategies involving the formal characteristics of pair members were used less frequently than verbal associative, imagery, and repetition (frequency) strategies, indicating that orthographic attributes are less important than the other three. Furthermore, imagery strategies were used predominantly with High-*I* pairs, while repetition was reported more often for Low-*I* pairs.

The results of these and other experiments isolate the imagery attribute as deserving further experimental attention, especially with a view to determining its relationship to situational frequency. Research on the effects of instructional sets in VDL (Rowe & Paivio, 1971a) has demonstrated that instructions to form a referential image to each of the R words in a list is superior to a repetition (frequency) instruction that requires *S* to repeat each R item aloud more often than the W item. This suggests that imagery may be a more effective variable than frequency, at least in some cases, but the relationship between the two can best be determined by contrasting them factorially in representative experimental situations. This was attempted in the present study.

The influence of a number of different word attributes on VDL is not necessarily incompatible with frequency theory, which is concerned with the effects of one particular attribute, situational frequency. It has been suggested, however, that even though other attributes may affect performance, "a frequency differential is a *dominant* [present author's italics] attribute in learning the usual VD task; when this attribute fails to serve as a discriminative cue, other attributes will become dominant (Underwood & Freund, 1970, [p. 285]." This position seems somewhat extreme. Why must it be required that the frequency attribute fail to serve a discriminative function before other attributes can be effective, or "dominant"? Or, from the present point of view, why must frequency be an ineffective attribute in the case of High-*I* words before imaginal coding mechanisms can come into play? It is just as reasonable to assume instead that frequency cues can be utilized in situations where other cues are inefficient or inappropriate, but, at the same time, these other cues may oftentimes be effective and indeed preferred even when discrimination on the basis of frequency is still possible.

The experiments reported here constitute an investigation of the roles of imagery and frequency attributes in VDL. These were manipulated via defining operations which are known to affect performance, viz., imagery instructions and the relative *I* value of the items in the case of imagery, and repetition instructions and within-list repetition of R and W items in the case of frequency. The relationship between the two was examined by comparing learning of High-*I* and Low-*I* pairs as a function of repetition instructions (Exp. I), by comparing repeated-W versus repeated-R item conditions as a function of imagery instructions (Exp. II), and by comparing repeated-W versus repeated-R conditions as a function of the *I* value of pairs (Exp. III).

EXPERIMENT I

The purpose of Exp. I was to determine whether *I* is an effective attribute in VDL even when use of a frequency cue is encouraged by *E*. The frequency cue was provided by having *S*s make extra overt pronunciations of the R items in a list of High-*I* and Low-*I* pairs. If frequency is the "dominant" cue in VDL, or at least if frequency predominates over imagery as a learning strategy, then the differential ease of learning High-*I* and Low-*I* pairs should be reduced under repetition instructions. Frequency theory would therefore predict an interaction between instruction and imagery level of pairs, such that the difference between High-*I* and Low-*I* pairs is less for the repetition group than for the control.

Method

Subjects.—The *S*s were 32 introductory psychology students at the University of Western Ontario who participated as part of a course requirement. None had served in previous VDL experiments. The *S*s were assigned in rotation to the Repetition and the Control conditions.

Lists.—Sixty-four, two- and three-syllable words four to eight letters in length were selected from the Paivio, Yuille, and Madigan (1968) norms. Half of the words had *I* values above 6.06 (High-*I*), and the remainder had *I* values below 3.61 (Low-*I*). All of the words had F values greater than 39 per million by the Thorndike-Lorge G count. Two sets of 16 High-*I* and 16 Low-*I* pairs were constructed by a random procedure, and these were combined to form two separate 32-pair mixed lists.

Procedure.—The *S*s were tested individually using a memory drum. A study-test procedure was employed, with the pairs being presented at a 5-sec. rate. The intertrial interval was 5 sec. On each study trial, one randomly selected member of each pair was underlined and *S* pronounced each word aloud. Pronunciation was required to ensure that *S* attended to both items. The *S*s in the Repetition condition were told to repeat the underlined word aloud three additional times (cf. Rowe & Paivio, 1971a) and that this should help them remember what the correct word was on the test trials. They were reminded to keep using this strategy at the beginning of Trials 2 and 3. Control *S*s received no mnemonic instructions. On the test trials, the underlining was absent and *S* chose the member of each pair that he thought was correct. Guessing was encouraged. One study-test trial on a practice list of four pairs, presented on flash cards, preceded the four trials on the experimental list. Four different random orders of the 32 pairs were used. Counterbalancing for the spatial position of R and W items was carried out such that for any given study or test trial subsequent to the initial presentation, item position for half of the pairs was changed from the preceding trial, but with the items of any given pair changing position an equal number of times (four) across all four orders. The R and W items occurred equally often in the left and right positions for each presentation of the list. Half of the *S*s in each group received one set of pairs, and the remainder the other set. Also, within each of these groups, half of the *S*s had one member of a pair correct, while for the remaining *S*s the other word was correct.

Results and Discussion

Two separate analyses of variance were performed on the data for each of the three reported experiments, one based on the number of correct responses and the other using the number correct minus the number of errors obtained by each *S*. The latter is the traditional correction for guessing and was used to obtain a relatively uncontaminated estimate of *S*'s discrimination ability. The *S*s were encouraged to guess when uncertain of the correct response, and the number of omission errors was less than .5% of the total responses in each experiment, so the correction seems justified. In any event, parallel results were obtained with the two sets of analyses, and only the corrected data will be reported here.

An analysis of variance with instruction, imagery, and trials as factors produced significant effects of instruction, F (1, 30) = 6.27, $p < .05$; imagery, F (1, 30) = 36.61, $p < .001$; and trials, F (3, 30) = 36.09, $p < .001$. None of the interactions were significant. Collapsed across trials, the results are presented in terms of proportion correct in Table 1. It is clear that while repetition facilitated overall performance, the effect of *I* was equally strong in both instruction groups. Since the imagery factor involved repeated measures, it was necessary to determine whether the data of Table 1 conformed to an additive model before calculating ω^2 (Vaughan & Corballis, 1969). Tukey's test (Myers, 1966, pp. 166–169) indicated that the additive model is applicable, F (1, 30) = 2.46, $p < .10$, and calculation of ω^2 showed that the instruction factor accounted for 41% of the total variance while imagery accounted for 32%.

An interaction between repetition instruction and imagery value of pairs in VDL, predicted on the basis of frequency theory, was not obtained. The results indicate that the effects of imagery and fre-

TABLE 1

Mean Proportion Correct in Verbal Discrimination Learning

Frequency cond.	Imagery cond.	
	High	Low
Exp. I		
Repetition instruction	.80	.66
Control	.63	.45
Exp. II		
SR	.96	.89
SW	.88	.72
Exp. III		
SR	.81	.58
SW	.50	.32

Note.—High and low imagery conditions refer to High-*I* and Low-*I* pairs in Exp. I and III, and to imagery instructions versus no instructions in Exp. II.

quency are independent, and therefore constitute evidence against a general dominance of frequency cues in VDL. While the frequency attribute may override other item characteristics in certain situations, this was clearly not so in the present case, despite the fact that *S*s were instructed to use differences in situational frequency as a response cue. On the other hand, the findings concur with previous results (e.g., Rowe & Paivio, 1971b) in showing the effectiveness of *I* as an attribute in VDL.

Experiment II

Experiment I showed that instructing *S*s to use a repetition (frequency) strategy in learning a VDL list did not affect the relative ease of learning High-*I* as opposed to Low-*I* pairs, which presumably reflects a differential use of imagery for the two pair types. In Exp. II, it was asked whether the spontaneous use of frequency cues as a response strategy would be affected by instructions to use imagery. The design involved the comparison of an imagery instruction with an uninstructed control condition as a function of the relative availability of frequency cues. Frequency was manipulated by the within-list repetition of either R or W items (Ekstrand et al., 1966). Repeating the R items in different pairs presumably enhances the use of frequency cues, whereas repetition of W items acts to offset or retard the frequency difference which normally occurs between the items of each pair.

Method

Subjects.—Sixty-four *S*s, selected from the same pool as in Exp. I, were assigned in rotation to the four conditions generated by the factorial combination of two levels of list and two levels of instruction.

Lists.—Two separate 32-pair lists were constructed by a random pairing procedure from a sample of 64 High-*I* words selected from the Paivio et al. (1968) norms. These items had the same characteristics as described for the High-*I* words in Exp. I. Each R item in the two lists occurred twice, each time paired with a different W item, to produce the same-right (SR) list condition. The two same-wrong (SW) lists were obtained by simply interchanging the roles of the R and W items in the SR lists, with each R item becoming a W item and vice versa.

Procedure.—The testing procedure was basically the same as in Exp. I with four study-test trials being given. The *S*s in the imagery instruction condition were given the single image instruction used by Rowe and Paivio (1971a), i.e., to form an image to the R word of each word on the study trials. Control *S*s received no mnemonic instructions.

Results and Discussion

An analysis of variance was performed on the number of correct responses (corrected for guessing) with list, instruction, and trials as factors. Significant main effects of list, $F\ (1,\ 60) = 9.84,\ p < .01$; instruction, $F\ (1,\ 60) = 8.30,\ p < .01$; and trials, $F\ (3,\ 180) = 61.55,\ p < .001$; were obtained, as well as significant interactions of Instruction × Trials, $F\ (3, 180) = 14.03$, $p < .001$, and List × Trials, $F\ (3,\ 180) = 5.15,\ p < .01$. The interactions reflect the fact that the largest increase in scores across trials occurred for the uninstructed SW group. This appears due to a general ceiling effect for the remaining three groups, whose scores reached asymptote on Trial 3.

The important finding from the present point of view was the absence of a List × Instruction interaction, indicating that the two factors are independent. The mean proportion correct for each condition is shown in Table 1, where the data are collapsed across trials. The list variable accounted for 13% of the variance in these data, and instruction accounted for 9%, with the remainder attributable to experimental error. The results, however, are not as clear-cut as those of Exp. I with respect to the relationship between imagery and list (frequency) effects. While the List × Instruction interaction did not approach statistical significance, the difference between Cond. SW and SR was smaller for the group which received imagery instructions. It could be argued that this disparity is an artifact of the overall ease of learning in the imagery instruction groups. In particular, the SR-Imagery *S*s made very few errors, an average of one or less on the final three trials. However, an analysis of variance of the first trial data failed to produce a significant List × Instruction interaction, although the two main effects were highly significant. In general, therefore, the results point to the conclusion that the effects of the two variables are independent. Further evidence for this conclusion is presented in Exp. III.

EXPERIMENT III

In this experiment, the roles of imagery and frequency in VDL were studied by comparing learning of High-*I* and Low-*I* pairs as a function of SR and SW list conditions. The experiment was intended to add to the generality of the results of the previous two experiments by introducing the imagery and frequency manipulations in different combination.

Method

Subjects.—The *S*s were 32 introductory psychology students, assigned in rotation to the two list conditions.

Lists.—The 32-pair mixed lists of High-*I* and Low-*I* pairs used in Exp. I were modified for use in this experiment. The two SR lists were formed by replacing the R items in half of the High-*I* and Low-*I* pairs with another R item of the same type from elsewhere in the list, with the restriction that each R item occur exactly twice in different pairs. The SW lists were obtained by reversing the correctness of the items in each pair of List SR.

Procedure.—The same procedure as used in the preceding experiments was followed here, except that the presentation rate was changed to 3 sec. in an attempt to increase the difficulty of the task and thus eliminate the occurrence of the ceiling effect observed in the data of Exp. II. Learning proceeded for four study-test trials.

Results and Discussion

The analysis of variance of the corrected data produced significant main effects of list, F (1, 30) = 18.07; imagery, F (1, 30) = 31.86; and trials, F (3, 90) = 40.12, all ps < .001. The List × Trials interaction was also marginally significant, F (3, 90) = 2.92, $p < .05$, the result of a somewhat sharper increase in correct responses over trials for the SW group. However, in no condition did the scores exhibit asymptotic performance. The List × Imagery interaction was clearly nonsignificant, as suggested by the data of Table 1, where the trials factor is eliminated. These data were shown to be additive by Tukey's test, F (1, 30) = 1.41, $p < .10$, and calculation of ω^2 showed that list accounted for 49% of the total variance while imagery accounted for 26%. The independence of imagery and frequency factors in VDL thus again was demonstrated in this experiment. These results were free from the ceiling effect associated with the data of the previous experiment and are regarded as strong evidence favoring an independence hypothesis.

DISCUSSION

Imagery and frequency, as defined by the present operations, appear to exert independent effects in VDL. The previously shown results also provide evidence that the two processes are additive. This may be seen most clearly from Table 1. Considering the proportions for the appropriate conditions in each experiment as estimates of the operation of imagery or frequency cues, it is possible to predict the level of performance when these occur jointly and to compare the predicted values with those actually obtained. Thus, in Exp. I, the estimated contribution of frequency alone is .66, and for imagery, .63. The addition law of probability predicts that the probability of occurrence of either or both together is .88. This value is somewhat higher than the .80 obtained for High-*I* pairs in the Repetition condition, where both cues are available, but does not negate the possibility that the effects are additive. Additivity is strongly suggested by the data of Exp. II and III, where the expected values (.99 and .79) are quite close to the obtained proportions of .96 and .81.

It should be mentioned that in each of the above experiments, *S*s were given an additional test trial following VDL, in which they were required to select from a list of four alternative strategies—repetition of the R items, imaging the R items, other, and none—the one which best described how they had learned each pair in the list. These data will not be presented in detail, but some general trends can be pointed out regarding the relative reported use of imagery and repetition. The two together accounted for about 70% of all strategies. The mean number of pairs for which imagery was reported was at least three times as great in the imagery conditions, both when the variable involved item attributes (Exp. I and III) and instructions (Exp. II), indicating the success of these manipulations in priming imaginal coding. For the frequency variable, however, the results were less clear. Repetition was reported about twice as often in the instruction group than in the control in Exp. I, but was unaffected by the SR-SW list conditions in the remaining experiments. Further questioning of *S*s in Exp. II and III revealed that two types of repetition strategies were contained in the repetition reports, one based on covert rehearsal of R items and the other based on the fact that items were repeated within the list. This breakdown of repetition showed that a repeated-items strategy was reportedly used more often for the SR lists, while rehearsal was used more often for the SW lists across both experiments, thus lending some support to the validity of the frequency manipulation. However, the repeated-items strategy was used in more than half of the cases involving SW lists, suggesting that many of the *S*s attempted to use a strategy based on repeated W items, despite the fact that this appears to have been an inefficient method of learning the list.

The independence of imagery and frequency as processes in VDL points to the conclusion that the imagery effect is not explainable in terms of frequency theory as presently formulated (cf. Rowe & Paivio, 1971a). Before offering an alternative explanation, a further specification of the nature of the imagery effect is necessary. The results of Exp. II above, along with those reported by Rowe and Paivio (1971a), demonstrate that a strategy which involves forming a referential image to the R item of each pair enhances VDL. On the other hand, an instruction to image to both words together with the R image designated as the larger of the two is no better than an uninstructed control condition (Rowe & Paivio, 1971a). This suggests that in order for an imagery mnemonic to work effectively in VDL, imaging should be restricted to only one of the items. The evidence involving *I* as the defining operation for imagery is also consistent with this position. Paivio and Rowe (1971) have shown that word pairs in which both members are High-*I* are learned no better than pairs with one High-*I* and one Low-*I* member, whereas both types are superior to Low-*I*–Low-*I* pairs. Providing the opportunity to image to both members of a pair does not enhance performance relative to the situation where only one item can be imaged, but the presence of at least one High-*I* word in each pair improves performance.

These results may be explained by assuming that a differential encoding response (DER) to the R or W item of a pair, in this case an image, can provide the basis for a correct discrimination. If *S*s consistently image to either the R or W member of each pair during study or feedback trials, this information can be used at the time of testing in deciding which word is correct. The DER produced by imagery in this way may be conceptualized as an associative "tag" which is stored with one of the words of the pair and is later re-elicited by the appropriate word to provide the cue for discrimination. This type of distinctive tagging is to be distinguished from the image-tagging notion previously proposed to account for superior learning of High-*I* pairs (Paivio & Rowe, 1970, 1971; Rowe & Paivio, 1971a), which assumed that *both* members of a pair are imaged and the *image* corresponding to the R item is tagged in some way. The evidence cited above favors the simpler interpretation involving direct associative tagging of only one item.

Analogous tagging mechanisms have been suggested by other authors. Kausler and Boka (1968) demonstrated that the way in which *S*s learn a double function list is by forming an R-W association for each pair and using the W items as discriminative tags to remember which words are correct. Distinctive tagging of R items has also been proposed to account for VDL under conditions of high intrapair similarity (Buschke & Lenon, 1969; Kausler & Kollasch, 1970; Kausler & Olson, 1969).

The theoretical distinction between discriminative tagging of items via DERs and frequency theory deserves some comment. One could, of course, view the higher accrual of frequency to R items as one type of differential encoding. However, a DER, as presently formulated, is seen as involving a transformation of the initial representation for an item into some qualitatively different memory code. In addition to an image, a recoded version of the nominal stimulus could be a verbal label for a picture, a natural language mediator constructed from a nonsense syllable, or a simple verbal associate, to name a few possibilities. A frequency "code," on the other hand, is best viewed as involving *quantitative* increments of the initial memorial representation of an item, which, in the case of verbal stimuli, is probably acoustic in nature. Rehearsal strategies, which affect differential frequency accural, and encoding strategies, which produce DERs, are probably both involved in a typical VDL task. The further isolation of DERs in addition to imagery and the teasing apart of these from frequency influences should contribute to a fuller understanding of the verbal discrimination process.

REFERENCES

BUSCHKE, H., & LENON, R. Encoding homophones and synonyms for verbal discrimination and recognition. *Psychonomic Science*, 1969, **14**, 269–270.

EKSTRAND, B. R., WALLACE, W. F., & UNDERWOOD, B. J. A frequency theory of verbal discrimination learning. *Psychological Review*, 1966, **73**, 566–578.

FULKERSON, F. E., & KAUSLER, D. H. Effects of intrapair and interpair bidirectional associates on verbal-discrimination learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 307–310.

KAUSLER, D. H., & BOKA, J. A. Effects of double functioning on verbal discrimination learning. *Journal of Experimental Psychology*, 1968, **76**, 558–567.

KAUSLER, D. H., & KOLLASCH, S. F. Bilingual verbal discrimination transfer: Implications for verbal discrimination theory. Paper presented at the meeting of the Psychonomic Society, San Antonio, November 1970.

KAUSLER, D. H., & OLSON, R. D. Homonyms as items in verbal discrimination learning and transfer. *Journal of Experimental Psychology*, 1969, **82**, 136–142.

LOVELACE, E. A. Verbal-discrimination learning: Varied familiarization of correct and incorrect items. *Canadian Journal of Psychology*, 1969, **23**, 227–232.

MYERS, J. L. *Fundamentals of experimental design.* Boston: Allyn & Bacon, 1966.

PAIVIO, A., & ROWE, E. J. Noun imagery, frequency, and meaningfulness in verbal discrimination. *Journal of Experimental Psychology*, 1970, **85**, 264–269.

PAIVIO, A., & ROWE, E. J. Intrapair imagery effects in verbal discrimination and incidental associative learning. *Canadian Journal of Psycology*, 1971, **25**, 302–312.

PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology* 1968, **76** (1, Pt. 2).

ROWE, E. J., & PAIVIO, A. Imagery and repetition instructions in verbal discrimination and incidental paired-associate learning. *Journal of verbal Learning and Verbal Behavior*, 1971, **10**, 668–672. (a)

ROWE, E. J., & PAIVIO, A. Word frequency and imagery effects in verbal discrimination learning. *Journal of Experimental Psychology*, 1971, **88**, 319–326. (b)

UNDERWOOD, B. J., & FREUND, J. S. Two tests of a theory of verbal-discrimination learning. *Canadian Journal of Psychology*, 1968, **22**, 96–104.

UNDERWOOD, B. J., & FREUND, J. S. Verbal-discrimination learning with varying numbers of right and wrong items. *American Journal of Psychology*, 1969, **82**, 198–202.

UNDERWOOD, B. J., & FREUND, J. S. Relative frequency judgments and verbal discrimination learning. *Journal of Experimental Psychology*, 1970, **83**, 279–285.

VAUGHAN, G. M., & CORBALLIS, M. C. Beyond tests of significance: Estimating strength of effects in selected ANOVA designs. *Psychological Bulletin*, 1969, **72**, 204–213.

(Received December 3, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 147–153

RECOGNITION AND RECALL IN SHORT-TERM MOTOR MEMORY[1]

PHILIP H. MARSHALL[2]

Texas Tech University

The purpose of this research was to test the general proposition of J. A. Adams' 1971 closed-loop theory of motor behavior, that the process underlying movement is a perceptual one whereby *S* recognizes correct and incorrect responses when they occur. The strategy of the research was to run parallel recognition and recall experiments and then ask if recognition and recall measures of motor memory are the same functions of the same variables. Increases in reinforcement positively affected both recognition and recall, and increases in retention interval negatively affected both recognition and recall. Other findings relevant to closed-loop theory were discussed.

The purpose of this research was to explore implications from a closed-loop theory of motor behavior (Adams, 1971). Heretofore, the major attempts that have been made to develop systematic accounts of motor behavior have reflected trends in the psychology of learning and have been related to S–R theory which derives from Thorndike (e.g., Bilodeau, 1966). According to open-loop conceptualizations of behavior (although they are seldom called that), the response is controlled by stimuli acting upon the system. In general, these theories say that if the eliciting stimuli are adequate, and if habit and motivation and the like are sufficient, the response will occur, otherwise not.

In contrast to the open-loop positions are closed-loop conceptualizations (e.g., Adams, 1971) which emphasize feedback, comparison processes, error detection, and correction. The operation of a closed-loop system first requires the initiation of a response. Response-produced feedback is then compared with a reference mechanism that defines the correct response. If a successful comparison is made, with the feedback from the current response matching the reference, the response is considered correct and is emitted. If a successful match is not made and *S* senses an error signal, *S* will act in response to the error signal by either inhibiting any response altogether or by giving a new error-correcting response. An open-loop system sees errors as the momentary dominance of noncriterion responses over the criterion response, but errors in a closed-loop system are central and are related to the patterning and flow of behavior.

According to Adams' (1971) theory, the response-generating mechanism is called the *memory* trace, and it governs the selection of responses to be subjected to the comparison process. Also specified by the theory is the development of an internal reference mechanism for movement. This reference mechanism is called the *perceptual* trace and is defined by the occurrences of response-produced stimuli during the learning of the criterion response. The strength of the perceptual trace grows as a positive function of the number of times the stimuli from which it is developed are experienced. Subsequent feedback from a response is compared with the reference, and discrepancies are recognized as errors to be corrected. The theory is centrally concerned with the feedback mechanisms that are involved in the learning and retention process, and this important role for feedback has been confirmed in two recent studies (Adams, Goetz, & Marshall, 1972; Adams, Marshall, & Goetz, 1972).

---

[1] This research was based on a dissertation submitted to the University of Illinois at Urbana-Champaign in partial fulfillment of the requirements for the PhD degree. The author is indebted to his committee members, Harold W. Hake, Charles Lewis, William E. Montague, and Keith G. Scott, with special thanks to Jack A. Adams, committee chairman, for his encouragement and comments. The research was supported in part by United States Public Health Service Grant MH-12022 to Jack A. Adams.

[2] Requests for reprints should be sent to Philip H. Marshall, Department of Psychology, Texas Tech University, Lubbock, Texas 79409.

The focus of the current research is on the comparison between recognition and recall in short-term motor memory (STMM). According to the theory just reviewed, there should be a close correspondence of the effects of particular experimental manipulations on recognition and recall in STMM. This is so because the closed-loop theory places primary emphasis on the ability of *S* to recognize a correct movement at the time of test, and it is this ability that is the basis for his response at the time of test, whether that test be recognition or recall. Indeed, though all the studies of STMM to date have reported the use of a "recall" procedure, the tasks actually required *S*s to reproduce a prior movement which closed-loop theory sees as essentially a recognition task in that *S* moves until he "feels" or recognizes that he is correct.[3] The hypothesized correspondence is *only for situations which do not allow the memory trace to operate*, and this condition has been satisfied by nearly all the STMM studies to date. An example is the simple linear positioning response which precludes the operation of the memory trace in that starting position, path, and direction of movement (elements which if free to vary would be dependent on the memory trace in that the function of the memory trace is to select the response) are all fixed, leaving room only for the perceptual trace to operate.

EXPERIMENT I

Method

Purpose.—The purpose of this experiment was to assess the role of recognition for STMM by asking if *S* can recognize a criterion movement when it is presented again by discriminating it from alternatives, and if short-term motor recognition memory has the same dependency on number of reinforcements and duration of retention interval as STMM measured by recall (Adams & Dijkstra, 1966). The method was a two-interval, forced-choice recognition test to determine *S*'s ability to discriminate movements longer or shorter than the criterion movement that was previously made.

Design.—Between-*S*s variables were two reinforcement conditions (one vs. six response repetitions), three durations of retention interval (5, 60, and 90 sec.), and four levels of discriminability of alternatives (±5%, ±10%, ±15%, and ±20% of criterion). Within-*S*s variables were testing interval (whether the criterion response was presented as the first or second alternative on the recognition test) and type of alternative (longer or shorter than the criterion movement).

Apparatus.—The apparatus consisted of a free-moving slide mounted on a stainless-steel track. Attached horizontally to the slide was a handle with which *S* moved the slide. On a second track mounted above *S*'s slide was a second slide equipped with a set screw. Positioning of this second slide determined how far *S*'s slide travelled from the starting position.

Subjects.—There were 10 right-handed, male *S*s in each of 24 groups. The *S*s were fulfilling a requirement for a course in introductory psychology at the University of Illinois. The task required a movement from right to left and for this reason only right-handed *S*s were used.

Procedure.—The procedure was to first present the criterion movement and then the two test movements, one of which was the same as the criterion movement. This was essentially a two-interval forced-choice testing procedure with the probability of the criterion length being presented in each interval being .50. The use of instructions designed to inform *S* of the random assignment of criterion length to each interval and to caution against the use of a definite response pattern (e.g., 1, 2, 1, 2, 1, 2, etc.) reduced the likelihood of response bias.

There were five different criterion lengths used (14, 18, 22, 26, and 30 cm.). Each length was presented with a long and short alternative, and either in the first or second test interval, resulting in four presentations of each length for a total of 20 trials per *S*. The presentations of combinations of criterion length, long or short alternative, and test interval were randomly determined for each *S*.

For all conditions, a 30-sec. intertrial interval (ITI) was used, measured from the end of *S*'s movement on the second test trial to the beginning of the next criterion movement.

The instructions described the task and emphasized the random nature of the presentations of the criterion length on the two test trials. In addition, the instructions included a practice trial with a 5-sec. retention interval at a length not used in the study (24 cm. with the alternative set at +15 cm.). Except during instructions, *S* was blindfolded throughout. The *S* was seated so that his right arm was directly in front of the slide handle at the starting position and so that his left arm rested on the other arm of the chair. When not making a movement, *S* was instructed to keep his right hand in his lap. On the command "Grip," *S* grasped the handle and moved the slide at a moderate, deliberate speed until the stop was contacted. After contact was made with the stop, *S* returned his hand to his lap, leaving the slide in place. The slide was returned to the starting position by *E* and, after the specified retention interval, the two test movements were administered by *E* again giving the command "Grip" for each test movement. No more than 3 sec. separated the two test movements. If six reinforcements were to be given, *S* made six repetitions of the criterion movement before the onset of the retention interval, with no more than 3 sec. separating successive movements. After the two tests, *S* made his choice by giving the interval aloud (e.g., "one") and, in addition, rated the confidence he had in his choice on a scale ranging from 1 (least confident) to 5 (most confident). Confidence ratings were taken to provide more information on the state of *S*'s perceptual trace for the criterion movement in relation to the two test movements. The confidence rating, it is assumed (Adams, 1971, p. 127), should vary as a function of the success of the match between response-produced stimuli and the perceptual trace. Correct designation of the criterion movement should yield a higher confidence rating than an incorrect designation.

---

[3] The term recall will hereafter be used to describe a procedure that requires the reproduction of a prior response.

Results

Response bias.—The basic measure of this study was the proportion of responses correct, P(C), for the various treatments. In order for this to be a valid measure, it is necessary to show that no differential response-related bias existed among the various groups. The number of times each *S* chose the first of the two test movements as being the same as the criterion movement was entered into an analysis of variance, with number of reinforcements, retention interval, and discriminability of alternatives as between-*S*s variables. No main effects or interactions were significant at the .05 level (all ps $> .20$), with the mean number of times the first movement was chosen varying from 8.80 to 11.00 out of 20. Thus, there was no significant differential response-related bias associated with the various treatments.

The proportion of correct responses for each of the four combinations of two alternatives and two test intervals were computed for each *S* based on five observations, one from each of the five lengths. These proportions were entered into an analysis of variance, with reinforcement, retention interval, and discriminability of alternatives as between-*S*s factors and alternatives and test interval as within-*S*s factors.[4] This analysis showed a significant effect of reinforcement, F (1, 216) = 11.74; retention interval, F (2, 216) = 6.10; and discriminability, F (3, 216) = 83.36, all ps $< .01$. Also significant in this analysis were effects due to alternatives, F (1, 216) = 6.05, $p < .02$; and the Alternatives × Test Interval interaction, F (1, 216) = 4.16, $p < .05$. No other effects were significant at the .05 level. Inspection of Fig. 1 shows that six reinforcements are superior to one reinforcement in virtually all conditions, that performance decreases as discriminability decreases, and that recognition performance tends to decrease as retention interval increases. The significant Alternatives × Test Interval interaction indicated that performance at the two intervals was essentially the same for the short alternatives but was superior at the first interval for the long alternatives. Recognition was greater for trials with long alternatives, P(C) = .72, than trials with short alternatives, P(C) = .67.

An analysis of variance was conducted on the five criterion lengths to determine if there was any difference in performance associated with the various lengths. The measure for each *S* was the proportion of the four responses correct for each length. A significant effect was found, F (4, 864) = 8.97, $p < .01$, with the mean proportion correct increasing with increases in the criterion length (.64, .64, .69, .73, .72, for each successive length, respectively).

Confidence ratings.—Ratings were incorporated in this study primarily because they may be used to assess the quality of the perceptual trace and response-produced stimuli match that occurs at the time of test. The *S*s should give a higher confidence rating for correct than for incorrect responses because a correct response is a successful match, and this success is revealed in a feeling of greater subjective confidence. To assess this theoretical implication for ratings, a mean rating was obtained for both correct and incorrect responses for each *S*. To be included in this analysis, *S* must have had at least one correct and one incorrect response. This requirement excluded only three *S*s from the analysis because of perfect performance. Between-*S*s factors were re-

---

[4] All analyses using proportions were repeated with arc-sine transformations, and the decisions about the variables were the same as for the proportions themselves.

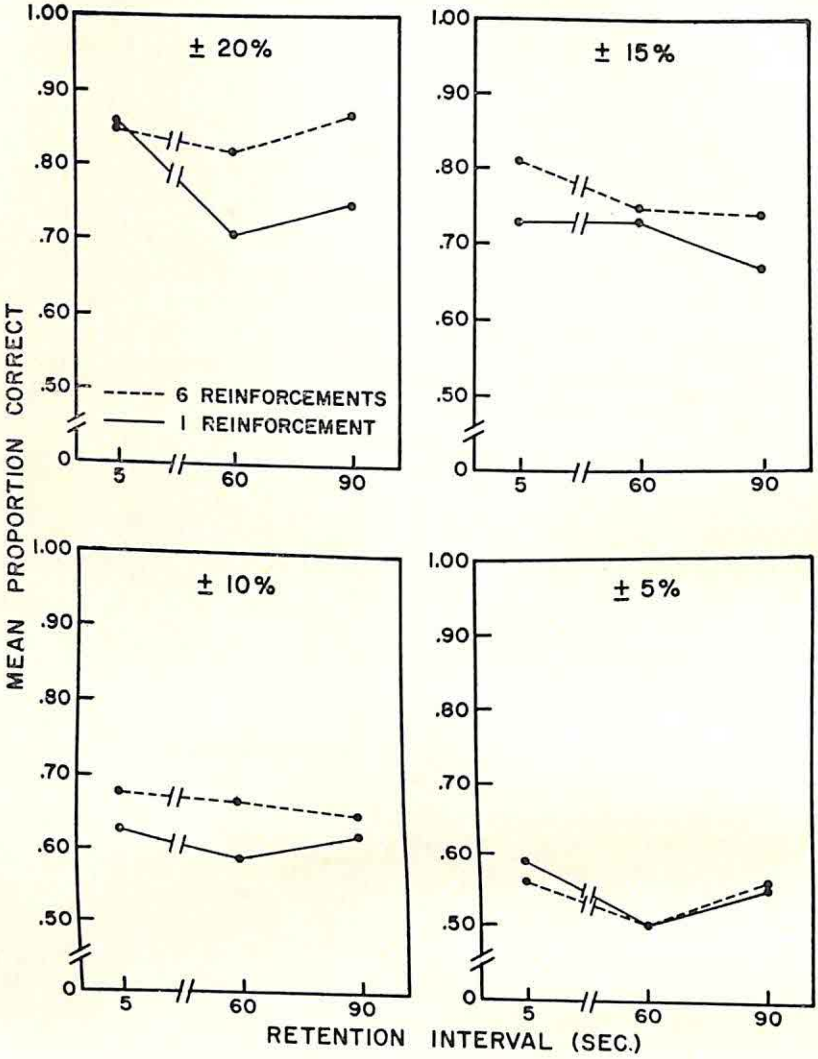


FIG. 1. Proportion correct as a function of reinforcements, retention interval, and discriminability of alternatives.

inforcement, retention interval, and discriminability, with correctness of the response as the within-*S*s factor. Of the between-*S*s factors, only discriminability was significant, $F\ (3, 213) = 11.80, p < .01$, with ratings increasing as discriminability increased (3.02, 3.36, 3.55, and 3.64). The effect of correctness was also significant with the mean rating for correct responses being 3.59, and for incorrect responses, 3.20, $F\ (1, 213) = 146.79, p < .01$. Also significant were the Reinforcements × Correctness interaction, $F\ (1, 213) = 6.58, p < .02$; and the Discriminability × Correctness interaction, $F\ (3, 213) = 8.64, p < .01$. Table 1 presents clarifying data for the Reinforcement × Correctness interaction and shows higher ratings for correct than incorrect responses, and the fact that

TABLE 1

MEAN CONFIDENCE RATING AS A FUNCTION OF NUMBER OF REINFORCEMENTS AND CORRECTNESS OF RESPONSE

Correctness	No. of reinforcements	
	1	6
Response correct	3.53	3.64
Response incorrect	3.22	3.17

TABLE 2

MEAN CONFIDENCE RATINGS FOR CORRECT AND INCORRECT RESPONSES AS A FUNCTION OF THE DISCRIMINABILITY OF ALTERNATIVES

Correctness	Discriminability			
	5	10	15	20
Response correct	3.13	3.48	3.78	3.95
Response incorrect	2.91	3.23	3.32	3.33

Note.—Discriminability is measured in percentages.

only for correct responses is there an appreciable increase in rating associated with an increase in reinforcements. Tables 2 presents data for the Discriminability × Correctness interaction and shows that with an increase in discriminability there is a rise and subsequent leveling off of the ratings for incorrect responses, whereas ratings for correct responses, at each point greater than incorrect responses, continue to rise with increased discriminability.

EXPERIMENT II

Method

Purpose.—The purpose of this experiment was to provide data for the comparison of recall measures and the recognition measures that were collected in Exp. I. Experiment II paralleled Exp. I in the essential variables and design.

Design.—Varied were the number of reinforcements (one vs. six) and the duration of retention interval (5, 60, and 90 sec.). These were between-*S*s factors, with 20 *S*s in each of the six groups.

Apparatus.—The same apparatus was used as in Exp. I.

Subjects.—One hundred twenty right-handed, male students were used. Of these students, 48 were recruited from an introductory course in psychology at the University of Illinois and were fulfilling a course requirement. The remainder (72) were paid $1.50 for their services. There were equal numbers of paid and nonpaid *S*s in each group.

Procedure.—The *S*s were randomly assigned to a reinforcement-retention interval condition, read the instructions describing the nature of the task, and given a practice learn-recall sequence at a length not used in the experiment (24 cm.). Upon the command "Grip," *S* grasped the handle of the slide and moved the slide at a moderate, deliberate speed until the stop was contacted. He then returned his hand to his lap, leaving the slide in place (*E* returned the slide to the starting position on all movements). If six reinforcements were given, this same procedure was repeated six times with no more than 3 sec. between repetitions. After the specified retention interval, *E* again said "Grip," and *S* moved the slide, attempting to reproduce the original movement, this time with the stop removed. Algebraic error was recorded in millimeters. Upon completing the test movement, *S* gave a rating of his confidence in the accuracy of his response on a scale ranging from 1 through 5, with 5 being most confident and 1 being least confident. The completion of the test movement marked the beginning of a 30-sec. ITI, which ended with *E* saying "Grip" for the learning of the next criterion movement. Each *S* received five such learn-recall sequences, with a different criterion length (14, 18, 22, 26, and 30 cm.) randomly assigned to each sequence for each *S*. The *S*s had vision during the instruction period but were blindfolded throughout the experimental trials.

Results

Absolute error.—Table 3 shows that absolute error decreased as reinforcement increased and increased with increases in retention interval. An analysis of variance found significance for the factors of reinforcement and retention interval, $F\ (1, 114) = 3.80$, $p = .054$; and $F\ (2, 114) = 6.78$, $p < .01$, respectively. The interaction between the two variables was not significant. These findings replicate those of an earlier study by Adams and Dijkstra (1966).

TABLE 3

MEAN ALGEBRAIC ERROR, ABSOLUTE ERROR, AND CONFIDENCE RATING AS A FUNCTION OF REINFORCEMENT AND RETENTION INTERVAL

Measure	No. of reinforcements and retention interval					
	1			6		
	5	60	90	5	60	90
Algebraic error	6.85	.07	−1.16	−2.93	.82	1.46
Absolute error	18.21	24.13	23.02	14.83	19.78	22.12
Confidence rating	3.74	3.60	3.64	3.84	3.34	3.66

Note.—Retention intervals are measured in seconds.

TABLE 4

Mean Algebraic Error, Absolute Error, and Confidence Rating for Criterion Lengths

Measure	Criterion length				
	14	18	22	26	30
Algebraic error	−3.69	−2.18	.24	.03	9.76
Absolute error	15.69	21.21	20.09	22.51	22.21
Confidence rating	3.79	3.68	3.62	3.50	3.60

Note.—Criterion lengths are measured in centimeters.

An analysis of variance conducted with the criterion lengths as a within-*S*s factor indicated a significant positive relationship between the increase in criterion length and amount of error, $F\ (4, 456) = 3.30,\ p < .02$. As shown in Table 4, the longer the criterion length, the longer the error. This, too, was found by Adams and Dijkstra (1966).

Algebraic error.—No significant effects were found for reinforcement or retention interval when the measure entered into analysis of variance was mean algebraic error for each *S*.

In an analysis with the criterion lengths as a within-*S*s factor, a significant effect was obtained, $F\ (4, 456) = 5.63,\ p < .01$, and, as shown in Table 4, an increasing positive directional bias is obtained as the criterion lengths increase. Also significant was the second-order interaction between reinforcement, retention interval, and lengths, $F\ (8, 456) = 2.72,\ p < .01$.

Confidence ratings.—An analysis of variance on the effects of reinforcement and retention interval on the mean confidence rating for each *S* gave significance for retention interval only, $F\ (2, 114) = 3.74$, $p < .05$, but inspection of Table 3 reveals a nonsystematic trend, in contrast to absolute error. Other evidence of the lack of relation between absolute error and confidence rating is the insignificant correlation of $-.11$ obtained between them ($p > .05$) when the measures for each of the 120 *S*s were mean absolute error and mean confidence rating.

Discussion

The purpose of the research was to test the general proposition of a closed-loop theory (Adams, 1971) that the process underlying movement is a perceptual one whereby *S* recognizes correct and incorrect responses when they occur. Through experience, the feedback stimuli associated with a correct response lay down a perceptual trace that is the internal reference for the correct movement and is the basis for the recognition act. The strategy of the research was to run parallel motor recognition and recall experiments and then ask if recognition and recall measures of motor memory are the same functions of the same variables. Short-term motor memory was chosen as the empirical frame of reference, and the common variables were amount of reinforcement and duration of the retention interval. The measure for recall was the degree of accuracy in the reproduction of the criterion response, and for recognition the measure was the proportion of responses correct in a choice situation following learning, where the discriminability of alternatives in the choice test was also a variable. To obtain the proportion-correct measure, *S* compared each of two alternatives against a perceptual trace of a criterion movement he had practiced earlier and then chose the alternative that best matched the criterion movement. For the recall measure, *S* attempted to reproduce the criterion movement which he had made earlier. Though the two measures require different behavior on the part of *S*, the closed-loop theory sees them both as being determined by the state of the perceptual trace at the time of test, and they both should be the same function of the same variables. The results tended to support the theoretical assumption.

Theoretically, the amount of practice, or reinforcement, is postively related to the strength of the perceptual trace, and both the accuracy of recognition and of recall increased as the number of reinforcements increased from one to six. Weakening of the perceptual trace as a function of forgetting processes occurrs in the retention interval, and both the recognition and recall measures had a decline in accuracy over the interval between learning and test. The effect for retention interval was clear for absolute error in recall, but the loss for the recognition measure was small and less convincing even though the drop was statistically significant.

There is at least one possible reason why the retention interval effect was less for recognition than for recall. The procedures for the recognition study required each *S* to experience each criterion length four times. Under the one-reinforcement condition *S*s actually had up to 4 presentations of each length, and under the six-reinforcements condition *S*s actually had up to 24 presentations of each length. These additional experiences with the criterion lengths could have greatly reduced the amount of forgetting. If *S*s were actually benefitting from these additional repetitions of the criterion lengths, performance should have increased as trials increased and should have increased more for the six-reinforcements condition than for the one-reinforcement condition. However, an analysis of the proportion correct as a function of increasing blocks of five trials did not yield a significant effect for either blocks or the Reinforcement × Blocks interaction ($p > .05$).

Despite the measure of success for predictions from the closed-loop theory, one should be cautious in uncritically accepting the theoretical proposition that motor recall is fundamentally a recognition process because both measures could be a correlate of some other process. As an example, the S–R view might prove valid, when recall and recognition are both measures of the same associative state but with different sensitivities (see Adams & Bray, 1970, pp. 387–388). Nevertheless, the results of the experiments encourage the hypothesis. As future experiments view motor behavior as a recognition process and find positive results, a supporting empirical network will emerge to strengthen the hypothesis and diminish rival ideas. This study was the first to view motor learning and memory in terms of recognition, and it represents a start toward the expression of a new point of view in motor learning.

REFERENCES

ADAMS, J. A. A closed-loop theory of motor learning. *Journal of Motor Behavior*, 1971, **3**, 111–149.

ADAMS, J. A., & BRAY, N. W. A closed-loop theory of paired-associate verbal learning. *Psychological Review*, 1970, **77**, 385–405.

ADAMS, J. A., & DIJKSTRA, S. Short-term memory for motor responses. *Journal of Experimental Psychology*, 1966, **71**, 314–318.

ADAMS, J. A., GOETZ, E. T., & MARSHALL, P. H. Response feedback and motor learning. *Journal of Experimental Psychology*, 1972, **92**, 391–397.

ADAMS, J. A., MARSHALL, P. H., & GOETZ, E. T. Response feedback and short-term motor retention. *Journal of Experimental Psychology*, 1972, **92**, 92–95.

BILODEAU, I. MCD. Information feedback. In E. A. Bilodeau (Ed.), *Acquisition of skill.* New York: Academic Press, 1966.

(Received December 6, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 154–158

EFFECT OF STIMULUS INTERVAL AND FOREPERIOD DURATION ON TEMPORAL SYNCHRONIZATION[1]

PAUL R. BEST,[2] JR., AND NEIL R. BARTLETT[3]

University of Arizona

Temporal synchronization with isochronous, auditory stimuli was investigated by having three *S*s finger-tap to clicks immediately following a random foreperiod. Eleven different stimulus intervals ranging from .1 sec. to 4.5 sec. were presented randomly. A four-factor (foreperiod, interval size, signal position, *S*s) analysis of variance on the logarithms of the variances of the responses about each signal and about the mean response indicated that foreperiod had no significant effect. Interval size and signal position were the most important factors. At short intervals response was random, but variance of course was small. With increasing interval size the variance increased, but diverged increasingly from randomness. Analysis of the constant error demonstrated that it too increased with increasing interval size.

The importance of accurate timing for skilled performance is widely accepted (Michon, 1967). Several investigators have examined timing behavior itself by observing *S*s' ability to synchronize motor movements with external stimuli (e.g., Bartlett & Bartlett, 1959; Dunlap, 1910; Michon, 1967; Woodrow, 1932). Their investigations were attempts to look at pure timing behavior with minimal confounding of other tasks.

The data of Dunlap (1910) and of Woodrow (1932) were collected before the advent of sophisticated recording apparatus. Both investigators discarded the initial reactions of *S*s as they were interested only in the responses after adaptation to the frequency of stimuli. They measured only a very few short trains of responses; apparatus limitations made data recording very time consuming. Bartlett and Bartlett (1959) also allowed *S* to "follow" the stimuli before responding and then recorded only one response. Again, due to mechanical limitations, they were not able to record trains of synchronization, but rather were limited to one record per trial. Recently, however, high-speed electronic counter and printing units have been developed which permit extensive data collection involving attempts to synchronize with trains of fast-paced events.

Michon (1967) made a detailed and informative comparison of synchronization behavior (*S* attempts to synchronize his responses with the stimuli) with continuation behavior (*S* attempts to continue responding at the same frequency as the stimuli which have been discontinued). Michon also discarded the first several responses of the *S*s.

These studies do not address themselves to the question of what processes are involved when a human *S* initially attempts to synchronize. Is a certain minimum amount of time necessary to adapt to the stimulus frequency or are a certain minimum number of stimuli necessary?

It should be possible to instruct *S* not to only respond to the first of a train of regularly appearing stimuli, as would be done in a simple reaction time (RT) task, but to respond also to the succeeding stimuli, attempting to synchronize as quickly as possible. In this way something of the process of arriving at synchrony, or at least subjective synchrony, should be observable.

In order to carry out a study like the one suggested above, it is necessary to provide certain conditions of certainty and uncertainty. It is not desirable to confound synchronizing performance with choice behavior. The situation desired is one in which *S* is alerted, but begins with time uncertainty and then moves into a condition of time certainty as the train of stimuli are delivered. This makes it imperative that the first signal come at a time that is somewhat uncertain. Therefore, a foreperiod (FP) must be interpolated between a warning signal and the first stimulus which is variable enough so that *S* cannot anticipate when the first stimulus will arrive.

---

[1] This paper is based on an MA thesis by the senior author which was partially assisted by Grant GB-3955 from the National Science Foundation to the University of Arizona.

[2] Requests for reprints should be sent to Paul R. Best, Department of Psychology, University of Arizona, Tucson, Arizona 85721.

[3] The authors would like to thank Louis A. Pfautsch who constructed most of the apparatus and Charles T. Rasmussen and Terry C. Daniel for their suggestions and assistance.

The FP literature is only indirectly pertinent to the present issue since it is directed at single or refractory reactions. In a synchronization task, *S* is acting under different instructions and a different set. He must be prepared to react not only once, but several times in succession at varying paces. The effect of FP may be entirely different under such circumstances.

In order to assess the effect of FP, it is necessary to have at least one "long" and one "short" FP. The available FP literature would indicate that a short FP should not be longer than approximately 2 or 3 sec. (e.g., Woodrow, 1932). The long FP, therefore, could reasonably be set at 6 or 7 sec. In order to keep the FPs variable, a range of short FPs and one of long FPs were used.

The present study is an attempt to answer the question: How does *S* prepare to synchronize? Of interest are the processes involved prior to the responses recorded in previous studies. Knowledge of these processes should help optimize skilled performance for which accurate timing is important.

METHOD

Subjects.—Two female and one male undergraduate students were used as *S*s. All three were given 15-hr. training at the experimental task with intervals not used in the study. The *S*s were paid for their time.

All three *S*s were given an audiometric screening and tested as having normal hearing.

Apparatus.—The stimuli were generated and the nterstimulus interval controlled by a Hewlett-Packard 3300A function generator through a set of Roberts 54–04 headphones. The stimuli were clearly audible clicks of approximately 10-msec. duration. The *S* attempted to synchronize with the series of stimuli by tapping his finger, with a thin piece of copper wire attached, on a piece of silver-plated aluminum. His responses were measured by a Hewlett-Packard 5223L electronic counter and recorded by a Hewlett-Packard 5050A digital recorder. The error of recording was well within 1.0 msec.

The series of stimuli were preceded by a variable FP which was signaled to *S* via a warning light (GE Miniature Bulb No. 328, 6 v.). The FP was determined by a Hunter Decade Interval timer Model 111-C, Series D. When this timer was switched on, the warning light flashed and the FP began. When the FP ended, the FP timer automatically triggered another timer of the same type. This timer determined for how long a period the series of stimuli would be presented to *S*.

The *S*'s room was removed by some distance from *E*'s room to prevent *S* from hearing the noise of the equipment. All voice communication was via an intercom system.

Procedure.—All *S*s were instructed and trained to respond quickly to the first signal following the FP and also to each succeeding signal, making an attempt to synchronize to the best of their ability as soon as possible. The FP was either short (1.5–3.5 sec.) or long (6.5–8.5 sec.). FPs in .1-sec. increments were selected at random within these ranges for a given trial. Eleven stimulus intervals were used: 4.50 sec., 2.50 sec., 1.50 sec., 1.00 sec., .80 sec., .60 sec., .50 sec., .40 sec., .30 sec., .20 sec., and .10 sec. A single trial consisted of one of these frequencies being preceded by one of the FPs and continuing long enough to record the responses of *S* to the first 10 signals. The *S* was not aware that only the first 10 responses were monitored; his task was to synchronize as long as he could hear the stimuli. Prior to the initiation of the FP, *S* was questioned over the intercom as to his readiness. One run consisted of three warm-up trials (at frequencies not utilized in the actual test trials) followed by the 11 test frequencies presented in random order. The FPs for trials in a given run were always within the same range, i.e., either short or long; precise duration for each was chosen from a random list constructed prior to each run. The intertrial interval was 2 min. In a given session, four runs were presented to *S*; the FP order over runs was ABBABAABABBABAAB, etc. One session lasted approximately 90 min. Each *S* sat for 1 session on a given day and for a total of 10 sessions. Since each session consisted of two runs of each FP range a total of 20 replications for each condition was carried out on each *S*.

RESULTS

The difference in time between *S*'s response tap and the more nearly contiguous stimulus was measured. The algebraic

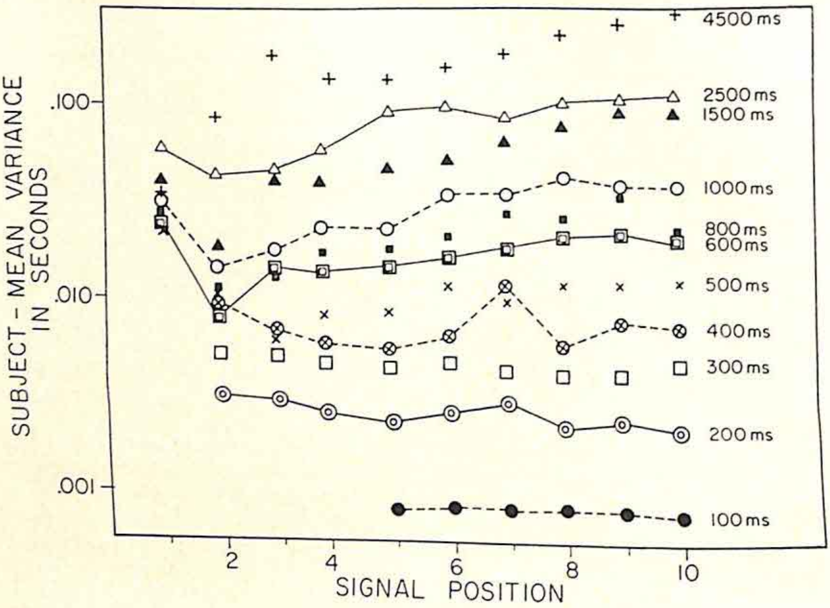


FIG. 1. Subject mean variance as a function of signal position.

sign was maintained with a negative value indicating an anticipatory response and a positive response indicating a lag. This difference measure constituted one error index. Since 20 replications were carried out, the 20 difference measures were combined into one variance measure. That is, the variance of the 20 responses about each stimulus in the train was calculated.

The variance of error about the signal is made up of two additive components. One of these is the variance within the distribution of errors, i.e., variance about *S*'s mean response (*S*-mean variance). The other component is the square of the constant error (CE^2).

A repeated-measures design was used with three factors (FP length, interval length, and signal position in the train) repeated for all three *S*s. Analyses of variance on the logarithms of the total variance and *S*-mean variance were calculated. The logarithmic transformation was chosen because variance measures were used as the dependent variable. Since variances are distributed as chi-square, the resulting distribution is positively skewed. The log transformation has the property of bringing such a distribution closer to normality for the analyses of variance. When means of logarithms are discussed it should be remembered that these are actually geometric means of the antilogarithms.

The signal position effect was significant for both variances, $F\ (9,\ 18) = 56.18$, $p < .01$; and $F\ (9,\ 18) = 151.43$, $p < .01$, respectively. Another significant contributor to the variability was interval length, $F\ (10, 20) = 98.38$, $p < .01$; and $F\ (10, 20) = 141.84$, $p < .01$, respectively. The two FP ranges did not produce significant differences. The Signal Position × Interval Length interaction was significant for both variance measures, $F\ (90,\ 180) = 8.66$, $p < .01$ for total variance; and $F\ (90, 180) = 8.91$, $p < .01$ for *S*-mean variance.

If *S* responds randomly to the fast intervals, he will produce a rectangular distribution of responses. The expected value of the variance of a random distribution throughout an interval is equal to the length of the interval in seconds squared, divided by 12. Therefore, the variances obtained experimentally can be compared with this expected variance for randomness (cf. Brownlee, 1965).

The data demonstrated a clear ordering of variance according to interval length. As interval length increases, variance increases accordingly. This is evident in Fig. 1. It appears that after two stimulus intervals, i.e., the third actual signal, consistent responding occurs for the 200- and 300-msec. intervals. Responses to the 100-msec. interval usually did not begin until about the fifth signal, but all thereafter are consistent though random. Intervals of from 400 to 800 msec. require at least three if not four stimuli for consistency. At 1,000 msec. and 2,500 msec. there is some leveling off apparent at the fifth or sixth signal while the remaining two intervals continue to increase the variability of the responses.

The foregoing statements are true for both total and *S*-mean variance. The means of the variances about Signal Positions 5–9 were calculated and plotted against log interval length in Fig. 2. Both ordinate and abscissa are log scales. The chance line is a plot of the log of random variance ($I^2/12$) across the log of interval length. The data for total variance are very similar. This graph illustrates the increase in log variance with log interval length and the increasing divergence from chance.

Interval length also produced an ordering of CE data, although not as marked as that for the variance data. In Fig. 3 is illustrated the fact that for most of the intervals *S* reached consistent levels of responding by the third stimulus. The longer the intervals the greater the negative CE produced, i.e., the greater the anticipatory responses.

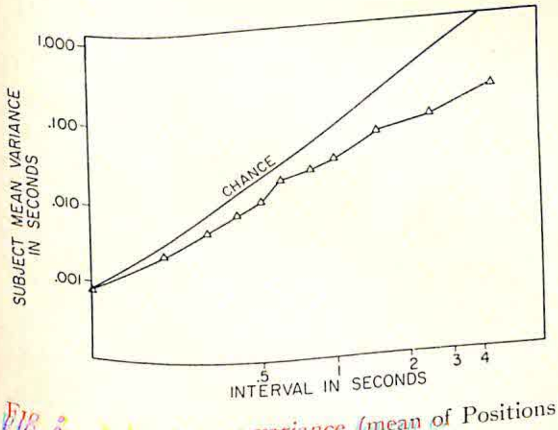


FIG. 2. Subject mean variance (mean of Positions 5–9) as a function of interval size.

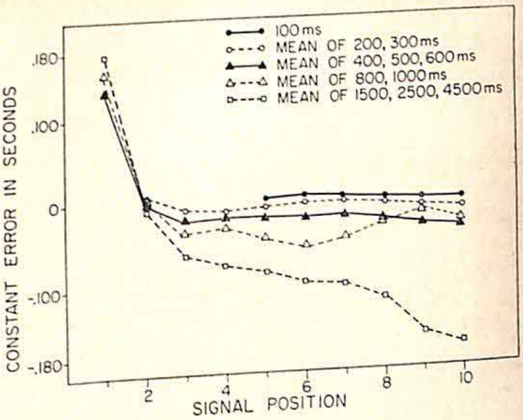


FIG. 3. Constant error as a function of signal position.

DISCUSSION

The two FP ranges used had no differential effect on the error variances. The lack of significance of this factor is not surprising as the distributions of FPs in the present study were random. Klemmer (1957) found that the variance in the distribution of FPs is a more important factor than the length. Klemmer's finding and the present finding are contrary to other studies, notably Woodrow (1914) and Woodworth (1938), which set forth 2.0 sec. as the optimum FP for a fast RT. The RT portion of the present study is obviously not a typical RT paradigm. This is why it was felt necessary to include FP length as a factor in the study rather than rely on findings that relate to RT. The resulting finding is that random presentation of FPs within a range has no effect and therefore synchronization tasks in which FP is not a pertinent factor may use only one such range.

Earlier the question was raised as to what conditions are important for the initial processes of synchronization. Examination of the variance data for each of the 10 signals suggests that the longer the stimulus intervals the more repetitions required for optimum responding. Thus, no constant number of stimuli or amount of time can be specified. The CE data, however, indicate that three stimulus intervals will reduce CE to a consistent minimum for intervals up to approximately 1.0 sec. The relationship of stimulus interval information to accuracy of synchronization is not a simple one, especially as the intervals become longer than 300 or 400 msec. It may take only two intervals to reduce variable error, but three to reduce constant error.

Bartlett and Bartlett (1959) found more variability for a response in which *S* had no time uncertainty than in a simple RT. We now find a further increase in variability when *S* continues making responses for which he has no external time uncertainty. Bartlett and Bartlett hypothesized that "synchronization errors show the variability in the central triggering and motor execution independent of any variability in the sensory processing, but compounded instead with the variability in anticipatory mechanism [p. 216]."

The effect of interval length on synchronization is described basically by the same theory set forth by Bartlett and Bartlett (1959). Rapid stimuli are responded to at random due to a limit on the speed of sensory registration and on motor frequency. The marked decrease in error variance for the intermediate intervals they reported was not found in the present study. The different methodologies utilized in the two studies may account for this. Whereas in the earlier study *S*s synchronized with one stimulus (of *S*'s choice, though usually the third or fourth), *S*s in the present study attempted to react to every stimulus, eventually beginning to synchronize usually by the third or fourth stimulus. Since they are not only attending to the first few stimuli but also are making motor responses, variability is increased, perhaps by a feedback mechanism as suggested by Michon (1967). Increased interval length increases the variance enough in the present study to overcome the reduction in variance found by Bartlett and Bartlett for their intermediate intervals. The longer intervals produce still larger variability. This is a result of *S*'s imperfect time memory, but there is no clear-cut delineation in the interval lengths as to where this memory element begins its effect.

One indication of the memory element and its effect on synchronization is in the ratio of CE to the square root of the *S*-mean variance (hereafter referred to as *S*-mean error). Following Chapanis' (1951) arguments about relative importance of variance components, these data can be interpreted as meaning that for the most part CE, for the present study, is a negligible part of the total variance; CE is considered negligible if its ratio to *S*-mean error is less than one-third. Thus, the variable error (*S*-mean error) is the major contributor to the total variance. In other words, the processes involved—anticipation, feedback, sensory registration, etc.,—contribute variability rather than a consistent bias. There are some exceptions to this general finding. The *S*-mean variance and CE increase across signal position more for the longer intervals than for the shorter ones. The Signal Position × Interval Length interactions for these two error measures indicate that the longer intervals produce an increase in variance across the train of stimuli while the shorter intervals do not.

The fact that CE has a greater proportionate influence in these particular position and interval combinations demonstrates the differences in responding to rapid rather than slow stimuli. While the stimuli are presented slowly enough, synchronization attempts are limited by *S*'s time memory which adds a consistent bias as well as more variability.

This study suggests, as did Bartlett and Bartlett's (1959) findings, that when an accurate timed response is required, it is best elicited by preceding it with two or three regularly appearing stimuli. If continued synchronization is required, stimuli separated by an interval in the range from .3 to .6 sec. should be utilized for optimum responding.

REFERENCES

BARTLETT, N. R., & BARTLETT, S. C. Synchronization of a motor response with an anticipated sensory event. *Psychological Review*, 1959, **66**, 203-218.

BROWNLEE, K. A. *Statistical theory and methodology in science and engineering*. New York: Wiley, 1965.

CHAPANIS, A. Theory and methods for analyzing errors in man-machine systems. *Annals of the New York Academy of Sciences*, 1951, **51**, 1179-1203.

DUNLAP, K. Reactions to rhythmic stimuli, with attempts to synchronize. *Psychological Review*, 1910, **17**, 399-416.

KLEMMER, E. T. Simple reaction time as a function of time uncertainty. *Journal of Experimental Psychology*, 1957, **54**, 195-200.

MICHON, J. A. *Timing in temporal tracking*. The Netherlands: Royal VanGorcum Ltd., Assen, 1967.

WOODROW, H. The measurement of attention. *Psychological Monographs*, 1914, **17** (5, Whole No. 76).

WOODROW, H. The effect of rate of sequence upon the accuracy of synchronization. *Journal of Experimental Psychology*, 1932, **15**, 357-379.

WOODWORTH, R. S. *Experimental psychology*. New York: Henry Holt and Co., 1938.

(Received December 8, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 159–163

CLUSTERING AND FREE RECALL WITH ALTERNATIVE ORGANIZATIONAL CUES[1]

RICHARD DOLINSKY

University of Toledo[2]

The present experiment investigated clustering and free recall when alternative clustering cues were available in the same word list. Specifically, the list composition allowed *S* to cluster items in associatively related pairs, rhyming pairs, or both. Associative clustering was chosen overwhelmingly. Further, though rhyming was used as a cue for clustering, the effect was clear only when there were no other nonsubjective cues. The incidence of rhymed clustering was considerably reduced when another, stronger, organizational basis was available.

The organized clustering of categorical items in free recall is, by now, a well-established phenomenon. Word associates, e.g., GIRL–BOY, tend to be recalled together at significantly higher-than-chance frequencies even though they were originally presented in a random order (Jenkins & Russell, 1952). On a nonsemantic level, Bousfield and Wicklund (1969) showed significant clustering in a list composed of 12 rhyming pairs; though Forrester and King (1971), due possibly to methodological differences, were unable to provide supporting evidence. In this last experiment, different words on the list could be organized in different ways, some semantically and others acoustically.

Bousfield and Wicklund (1969) used an unmixed list composed only of rhyming pairs; Forrester and King (1971), on the other hand, presented *S*s a mixed list consisting of semantically related, rhyming, and unrelated items. Further, presentation was oral in the Bousfield and Wicklund study and visual in Forrester and King. Although it would seem likely that oral presentation of items would maximize the possibility for acoustic clustering, it is known that auditorially similar materials can affect short-term memory even when presented visually (Conrad, 1964; Gumenik, 1969).

In the mixed-list design used by Forrester and King (1971), it is possible that the presence of semantically related items on the list focused attention, or imposed a set, toward them and away from the acoustically similar features of the rhyming words. The fact that overall recall of the rhyming items was not significantly different from recall of a group of unrelated words is consistent with this view. It should be reiterated that the semantically similar and rhyming items were members of different and exclusive sets, e.g., STEEL, IRON, and GOLD versus KEY, TEA, and SHE; that is, a particular item could be clustered in only one category. Nevertheless, clustering of rhyming items seems, at best, to reflect only weak organizational cues; even in the Bousfield and Wicklund (1969) experiment, the amount, though significant, was small.

It was the purpose of the present experiment to further investigate clustering in the recall of a list when alternative organizational modes were available for the same words. Specifically, the list composition allowed *S*s to cluster items in associatively related pairs, rhyming pairs, or both. Additionally, control lists were included in which only one type of categorical cue was present, either associative or rhyming. The use of these controls made it possible to more clearly evaluate the effects introduced into a list when two clustering alternatives were available.

---

[1] The author expresses his thanks to Kaisa Puhakka who assisted in the data collection and analysis of results for this experiment.

[2] Requests for reprints should be sent to Richard Dolinsky, Department of Psychology, University of Toledo, Toledo, Ohio 43606.

METHOD

Subjects.—The *S*s were 90 introductory psychology students from the University of Toledo. They were divided into three groups with 30 *S*s in each group.

Design and materials.—The *S*s in each group learned a 27-item (all common nouns and verbs) list in which 9 words were the same for all groups. These 9 "base" words were combined with 18 others in the following manner: in Group A-R (Associative-Rhyming), each of the 9 base words was accompanied by an acoustically dissimilar word associate and a rhyming nonassociate. For example, one of these 9, 3-word sets included the base word, BUY, its associate, SELL, and the rhyme, LIE. It was possible, then, for *S*s to cluster by association, rhyme, or both. For Group A-U (Associative-Unrelated), each of the base words was accompanied by the same word associates presented to Group A-R; but the rhyming items were replaced with 9 other words, unrelated both associatively and acoustically to the other items on the list. Consequently, clustering could occur only by association since there were no pairs that rhymed. Finally, in Group R-U (Rhyming-Unrelated), each of the base words was accompanied by the same rhyming items presented to Group A-R. The same 9 unrelated items from Group A-U made up the remaining words on the list. Consequently, there were only rhyming pairs on this list and no items related associatively.

Mean word frequencies for all conditions, i.e., associative, rhyming, and unrelated, did not differ significantly. The overall mean word frequency was 84.6 per 1.014 million as tabulated by Kučera and Francis (1967). Essentially, the word-associate pairs were bidirectionally related with a mean directional difference of only 12.6% (cf. Keppel & Strand, 1970).

Procedure.—Each list was tape-recorded and presented for three trials at a 3-sec. rate per word. A 90-sec. written recall period followed each trial presentation. Using a procedure similar to that of Bousfield and Wicklund (1969), on the first trial each word was both pronounced and spelled; on the remaining two trials the words were pronounced only. A different random order was used for the three-trial presentations with the restriction that no two items from the same category, either associative or rhyming, appeared in immediately adjacent presentation positions. The *S*s were run in subgroups composed of approximately 10 individuals.

TABLE 1

MEAN WORDS RECALLED AS A FUNCTION OF TRIALS FOR THREE GROUPS

Group	Trial		
	1	2	3
A-R	15.40	18.47	21.10
R-U	12.10	15.17	18.10
A-U	13.97	17.70	19.27

TABLE 2

MEAN CLUSTERED ASSOCIATIVE AND RHYMING PAIRS AS A FUNCTION OF TRIALS FOR THREE GROUPS

Cond.	Trial		
	1	2	3
Associative Clustering			
Group A-R	4.100	4.700	6.133
Group A-U	3.233	5.267	5.100
Rhymed Clustering			
Group A-R	.500	.900	.867
Group R-U	.833	1.633	2.367

RESULTS

Free recall.—The mean number of words correctly recalled is presented in Table 1. It is apparent that Group A-R produced the best learning with Groups A-U and R-U following in that order. The main effects of both groups and trials were significant at the .001 level, F (2, 87) = 10.13 and F (2, 174) = 158.87, respectively. The Groups × Trials interaction was not significant ($F < 1$).

The diverse list composition among the groups makes any conclusions somewhat speculative though there were nine words in common for all groups, the base words, and recall of these items was compared. The mean number of base words recalled correctly when pooled over trials was 7.07, 6.44, and 5.99 for Groups A-R, A-U, and R-U, respectively. The results of this analysis indicated a significant difference at the .01 level, F (2, 87) = 7.36; individual comparisons localized this effect to the superior recall in Group A-R over both Groups A-U and R-U. These latter two groups did not differ significantly. It is well known that recall of list items is enhanced if they appear with other related words. Apparently, the more bases of relatedness (association and rhyme, for example) the greater the probability for correct recall.

Clustering.—The mean number of pairs recalled in clusters is presented in Table 2.

Separate comparisons were made for word-associate and rhyming pairs. It should be noted that, in Group A-R, both clustering cues could be used effectively in the same recall; e.g., *S* could recall VIEW-SHOE-FOOT (or the reverse), though this type of organization was so rarely chosen that it hardly contributes as a relevant factor.

If we first consider the differences between associative and rhymed clustering in Group A-R, it is obvious that word association is an overwhelmingly preferred method, $F\ (1, 28) = 655.95$, $p < .01$. Comparisons of trials and Trials × Method of Organization were also significant, $F\ (2, 146) = 5.94$, $p < .01$ and $F\ (2, 146) = 3.38$, $p < .05$, respectively. The significant interaction reflects an increase in clustering over trials for the word associates but not the rhymed pairs.

The enormity of the difference between associate and rhymed clustering could, of course, stem from a number of factors. If the associates were weaker or if the rhyming pairs were different, the effect might be reduced. More specific analyses were computed between the same items in different groups in an attempt to further assess the role of associations and rhyming as organizational cues. Comparing the number of word-associated pairs in Groups A-R and A-U, results of the analysis of variance indicated no significant difference overall ($F < 1$), though trials and the Groups × Trials interaction were significant, $F\ (2, 116) = 19.88$, $p < .01$ and $F\ (2, 116) = 3.89$, $p < .05$, respectively. The interaction effect is difficult to interpret with more clustering occurring on Trials 1 and 3 in Group A-R. Although this group may allow an earlier start for clustering of associative pairs than Group A-U, it might be more conservative to conclude that whenever associative clustering is available to *S*, it will be used. Further, associative clustering is a robust factor since, by Trial 3, an average over both groups of 5.62 pairs (where 9 is the maximum) is recalled in this manner.

Similar comparisons of the organization of rhyming pairs yield a different picture. First, the incidence of clustering by rhyme, though present, is quite low. Again, with a maximum of 9 pairs possible, the best showing is on Trial 3 of Group R-U where the mean reaches only 2.37. These reliable, but small, occurrences of clustering are consistent with the results of Bousfield and Wicklund (1969).

Second, clustering differences with rhymes existed between Groups A-R and R-U, $F\ (1, 58) = 12.52$, $p < .01$; trials, $F\ (2, 116) = 15.10$, $p < .01$; and Groups × Trials, $F\ (2, 116) = 5.84$, $p < .01$. Specifically, clustering by rhyme was greater in Group R-U than in Group A-R. The poorer showing for the latter group suggests that when a stronger alternate method for organization is available, e.g., word association, *Ss* will use that method at the expense of weaker organizational cues; clustering by rhyme, therefore, suffers. Analysis of simple effects revealed an increase in rhymed clustering over trials only for Group R-U, $F\ (2, 116) = 19.58$, $p < .01$; a similar analysis for Group A-R was not significant. Consistent with the other findings, then, it appears that clustering by rhyme develops only when there is no strong alternative.

An attempt was made to assess the expected values for clustering along the lines suggested by Bousfield and Bousfield (1966) though, due to the presence of unrelated items on the list, this analysis is only tentative. Using the Bousfield and Bousfield formula for expected stimulus category repetitions, E(SCR), pairs were considered as categories consisting of two members and unrelated words as one-member categories.

For example, in Group R-U, if all 18 items making up the nine rhyming pairs (and only these items) are recalled, the E(SCR) would be 1.00; as additional unrelated items are also recalled, the maximum E(SCR) value would drop such that if all 27 list words are recalled, E(SCR) would be .67. Consequently, the E(SCR) values are only general estimates and they probably somewhat underestimate the chance amount of organization.

In Group R-U, all items recalled per trial were used in computing E(SCR) although, as noted above, the unrelated items were considered as one-member categories. Values of E(SCR) for Trials 1–3 were .46, .50, and .51, respectively. The differences between observed and expected clustering were significant by *t* test in all cases at the .01 level indicating reliable clustering by rhyme. For Group A-R, a similar attempt was made to estimate E(SCR) in regard to rhyme. Rhymed items were considered as two-member categories and, for computation purposes, the unrhymed, associate items were treated like the unrelated items in Group R-U. Values of E(SCR) for Group A-R were .27, .40, and .44 for Trials 1–3; observed minus expected clustering was significant on Trials 2 and 3 ($p < .01$ and $p < .02$, respectively), though the difference on the first trial was not reliable, t (29) = 1.49. Considering the earlier finding that rhymed clustering does not increase significantly over trials in Group A-R, the results of the observed minus expected comparisons might best be interpreted as providing only scant evidence for the development of any reliable clustering by rhyme. Clearly, for Trial 1, the observed number of pairings does not exceed chance expectations.

Clustering by word associates was strikingly significant over all trials in Groups A-R and A-U. Differences between various associative pairs were small, though different rhyming pairs showed occasionally quite variable effects. For example, SHOE and VIEW occurred together only 11 times while CHAIR and SHARE were clustered in 23 instances. This considerable variation was also found by Bousfield and Wicklund (1969), though the present study used totally different pairs.

DISCUSSION

There are several noteworthy findings in the present experiment. First, of course, is the major effect of word association over rhyming as an organizational basis in a list where either or both types are possible. Where a list is composed of items which allow alternative forms of organization, i.e,. Group A-R, the effectiveness of clustering by rhyme is considerably reduced. Though the present procedures are quite different from those used by Forrester and King (1971), these last results are in keeping with theirs. When cautionary evaluations are made between the observed and expected incidence of clustering in Group A-R, it appears that organization by rhyme may not much exceed chance expectancies when word-association cues are also present. Certainly, rhyme is not a major organizational cue in the first-trial free recall of Group A-R. Consistent with the results of Bousfield and Wicklund (1969), however, clustering of rhyming pairs in free recall does occur significantly, though this finding appears only in word lists where rhyme is the only apparent *E*-dictated mode of organization, i.e., Group R-U.

Though this study deals only with word association and rhyming, it seems likely that the issue exposed relates more to the relative influences of strong, e.g., association, and weak, e.g., rhyming, organizational cues present in the same list. Theoretically, *S*s in Group A–R could have used both cues for clustering though there was an observed decrease in clustering with weaker cues relative to a control situation (Group R-U) in which obvious competing forms of organization did not exist.

Why the rhyming cues are weaker in Group A-R than in Group R-U is not clear. Possibly if a base word is recalled in an associative cluster, only one slot is left open to cluster it with a rhymed item. That is, if a pair is recalled in the order, SELL–BUY, the possibility of additional rhymed clustering is restricted to the order, BUY–LIE. The recall slot immediately before BUY has already been occupied by the associate, SELL. On the other hand, in Group R-U, LIE could cluster with BUY in two orders, LIE–BUY, or the reverse. The result would be more rhymed clustering in Group R-U than Group A-R.

While this interpretation is a possibility, there are, however, at least two arguments against it. First, the present results are consistent with those of Forrester and King (1971) where alternative organizations of the same items were not available. Second, if the order for rhymed clustering in recall is restricted in Group A-R, a similar restriction would affect clustering by association. That is, if a pair is recalled in the order, LIE–BUY, the only slot left for associative clustering is the one immediately following BUY. If this is so, associative clustering in Group A-R should be less than in Group A-U where clustering could occur in either order, SELL–BUY or the reverse. The present findings, however, show no difference in associative clustering between the two groups ($F < 1$).

Essentially, the presence of alternative but weaker rhyming cues in Group A-R did not significantly effect clustering levels in the word-associate pairs where organizational cues were considerably stronger. Apparently, where one organizational mode is strong, the presence of alternative weaker types does not create measurable interference.

Overall free recall was greater in the list where both associative and rhyming cues were available; though it should be noted that availability of different alternative organizations is confounded with the specific words presented in each of the three groups. However, investigation of the words appearing in common on each list, i.e., the base words, allows a means of comparison. These results indicate greater free recall in Group A-R with no difference between the other two groups. This leads to the speculation that the more organizational formats a word can share, the greater the likelihood for correct recall. The base words, aside from their occurrence on the stimulus list for all groups, also are related both by association and rhyme to other words presented in Group A-R. For example, the base word, BUY, might be primed either by the list word, LIE, or by another list word, SELL. Or alternatively, it could be primed by both LIE and SELL together resulting in an increased probability for recall of the base word, BUY. That is, if the same words share various category organizations, rather than one, free recall might be facilitated. However, the relation between free recall and clustering in this situation is not at all clear.

In summary, the present experiment confirms earlier findings of significant but small amounts of clustering in rhyming pairs, and, even then, only when it is the only nonsubjective form of organization available. It also confirms, using a different experimental design, the finding of minimal, basically chance, clustering of rhyming items when the list provides other modes of organization. With the present materials, clustering by word association is overwhelmingly preferred. And further, it indicates the need for investigation into the conditions of organized recall when alternative cues are available. It seems possible, for free recall, that the facilitory role found in the present study when associative and rhyming cues are present in the same list might be less facilitory, and even interfering, if other organizational combinations are used.

REFERENCES

BOUSFIELD, A. K., & BOUSFIELD, W. A. Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports*, 1966, **19**, 935–942.

BOUSFIELD, W. A., & WICKLUND, D. A. Rhyme as a determinant of clustering. *Psychonomic Science*, 1969, **16**, 183–184.

CONRAD, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75–84.

FORRESTER, W. E., & KING, D. J. Effects of semantic and acoustic relatedness on free recall and clustering. *Journal of Experimental Psychology*, 1971, **88**, 16–19.

GUMENIK, W. E. Effects of articulatory activity and auditory, visual, and semantic similarity on the short-term memory of visually presented paired associates. *Journal of Experimental Psychology*, 1969, **82**, 70–74.

JENKINS, J. J., & RUSSELL, W. A. Associative clustering during recall. *Journal of Abnormal and Social Psychology*, 1952, **47**, 818–821.

KEPPEL, G. S., & STRAND, B. Z. Free-association responses to the primary responses and other responses selected from the Palermo-Jenkins norms. In L. Postman & G. Keppel (Eds.), *Norms of word association*. New York and London: Academic Press, 1970.

KUČERA, H., & FRANCIS, W. N. *Computational analysis of present-day American English*. Providence, R. I.: Brown University Press, 1967.

(Received December 13, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 164–170

EFFECTS OF REDUNDANCY LEVEL AND PRESENTATION METHOD ON THE PAIRED-ASSOCIATE LEARNING OF EDUCABLE RETARDATES, THIRD GRADERS, AND EIGHTH GRADERS[1]

HERMAN H. SPITZ[2]

E. R. Johnstone Training and Research Center, Bordentown, New Jersey

Retardates, normals of approximately equal chronological age (CA) (eighth graders) and normals of equal mental age (MA) (third graders) were presented with sets of paired associates which differed in redundancy level by varying the number of repetitions in the response terms. In Exp. I, presentation and testing were by the standard anticipation method. In Exp. II, presentation was simultaneous, and during testing all stimulus terms were presented simultaneously for free-ordered recall of the response terms. To achieve near perfect performance, the lower MA groups required more redundancy than the higher MA group. Compared to the anticipation method, and particularly on the low redundancy material, simultaneous presentation and free-ordered recall negatively affected the performance of retardates, positively affected the performance of eighth graders, and had no effect on third graders.

When all available information must be used to perform a task, then the information contains no redundancy. If, on the other hand, *S* requires less information than is available, the excess information can be considered redundant. Redundancy can be manipulated. For example, if six digits to be recalled are 1 3 5 1 3 5 the task is, theoretically, 50% redundant: the first three digits give all the information necessary to recall all six digits.

When ditits are 50% redundant, recall performance of educable mental retardates indicates that they do not readily recognize the presence of redundancy on initial presentation (Spitz, Goettler, & Webreck, 1972). However, when the redundancy is made salient, their performance improves markedly. With repeated presentations of digits which are 50% redundant, retardates discover and utilize redundancy, as reflected in improved performance (Spitz & Webreck, in press).

In paired-associate learning, there is evidence (e.g., Baumeister & Kellas, 1971) that retardates do not make effective use of secondary, redundant cues which are contained in the stimulus terms. Retardates behave more like normals only when their attention is directed to the relevant stimulus dimension. Brown (1970) has shown that in an oddity learning task, increasing the number of identical (redundant) cues improves retardates' performance to the level of equal chronological age (CA) normals who were not given the redundant cues. Making the odd cue redundant (e.g., by making it odd with respect to color as well as form) had the same facilitative effect.

In sum, there is evidence that in learning tasks, but not in certain perceptual recognition tasks (Spitz & Borland, 1971), level of redundancy and mental retardation interact. When redundancy is low or unemphasized, retardates have particular difficulty in recognizing and utilizing it. However, when redundancy is high or made salient, and/or when retardates have sufficient experience with a particular type and level of redundancy, their performance improves considerably.

---

[1] For assistance in obtaining normal *S*s, I am most grateful to Samuel Kaltovich, Principal of Roebling Public School No. 5, and to Sister Patricia, Principal of St. Raphael's School in Hamilton Township, New Jersey. I thank Barbara Swiderski for preparing the stimuli, and members of the Johnstone Research Department for critical reading of the manuscript. I am indebted to John J. Winters, Jr., for alerting me to the free-ordered recall method, and for suggesting that order of recall be recorded.

[2] Requests for reprints should be sent to Herman H. Spitz, Director of Research, Edward R. Johnstone Training and Research Center, Bordentown, New Jersey 08505.

It would be of value to find, on particular learning tasks, the level of redundancy which separates the performance of retardates from normals. To accomplish this, a paired-associate paradigm was adapted from Bower (1962). It permits varying the level of redundancy by varying the number of response alternatives. For example, if the S–R pairs are 6 B, 7 B, 3 B, 5 B, 8 B, and 4 A the response alternatives are highly constrained. The *S* merely has to recognize that 4 is paired with A and all other numbers are paired with B. In this example, the list is 83% redundant ($\frac{5}{6}$ of the responses are redundant) and contains two units of information (unit being used here as a chunk or item of information rather than as a traditional "bit"). That is, *S* has to learn that 4 is paired with A (one unit) and all the rest of the numbers are paired with B. If we change an additional response term from B to A, the list becomes 67% redundant and contains three units of information.

EXPERIMENT I

Method

Stimuli.—There were six paired associates at each of four redundancy levels, presented here in the order given in the single study trial: Redundancy Level 33: 8 C, 3 B, 6 B, 7 A, 5 C, 4 A; Redundancy Level 50: 7 A, 4 A, 8 B, 6 B, 3 B, 5 A; Redundancy Level 67: 6 B, 7 A, 4 A, 3 B, 5 B, 8 B; Redundancy Level 83: 6 B, 7 B, 4 A, 8 B, 3 B, 5 B. The use of numbers and letters reduced the possibility of attributing group performance differences to stimulus or response familiarity.

Subjects.—There were a total of 144 *S*s in Exp. I. The 48 retardates (16 girls, 32 boys; $\overline{CA}$ = 15.30, SD = 1.55; $\overline{IQ}$ = 64, SD = 6) were from a state residential institution (Johnstone). The 48 third graders (24 girls, 24 boys; $\overline{CA}$ = 9.16, SD = .39; $\overline{IQ}$ = 107, SD = 9) and the 48 eighth graders (16 girls, 32 boys, $\overline{CA}$ = 14.22, SD = .43, $\overline{IQ}$ = 107, SD = 9) were from a neighboring public school system. The IQs for the normals were from a group-administered California Test of Mental Maturity; for the retardates IQs were from individually administered Wechsler Scales. In each group there were 12 *S*s in each of the four redundancy levels. Male–female ratios were equated and IQ and CA were approximately equated within each *S* group across redundancy levels.

Procedure.—The *S*s, tested individually, sat approximately 4 ft. from a screen onto which the stimuli were back projected. Lower mental age (MA) *S*s were pretested for reading of numbers and letters. If they passed the pretest, they were given instructions which included a sample task in which an S–R pair (2 ×) was displayed on an index card, followed immediately by the stimulus alone. A correct verbal response (×) assured that *S* understood the task.

Each *S* then received one study trial which consisted of the six paired associates at his redundancy level. Each pair was exposed on the screen, one pair at a time, at a 3-sec. rate, while *S* read them aloud. (After the initial study trial, *S*s were not required to read the pairs aloud.) This was followed approximately 7.5 sec. later by the first anticipation trial block, in which the stimulus alone was presented for 3 sec. followed approximately .8 sec. later (Carousel change time) by a 3-sec. exposure of the S–R pair, and so on for the six pairs of a trial block. The order of the pairs was randomized over the first 5 trial blocks, at the end of which—if criterion had not been reached—*S* started over again with the pairs from the first trial block, and so on up to 30 trial blocks. The same order was maintained for all *S*s in each redundancy level. Trial block intervals were 7.5 sec. The criterion for learning was 2 consecutive correct trial blocks. Testing terminated if this criterion was not reached by the end of the thirtieth trial block.

Results and Discussion

There were no sex differences within any of the groups. Analyses could be made using trials to criterion or using errors. Since nine retardates (four at 33%, three at 50%, and two at 67% redundancy) and three third graders (one at 33% and two at 50% redundancy) never reached criterion, it was felt that an errors analysis would more truly reflect the data. Consequently, percentage of error out of a possible 180 paired associates (six pairs per trial block, 30 trial blocks) was used for all analyses. This measure assumes that *S*s who reached criterion would score correctly on all trials beyond the criterion.

Results are shown in Fig. 1, where it can be seen that with one exception, there was a steady increase in errors as redundancy level decreased. Analysis of variance indicated that group and redundancy level were significant main effects ($p < .005$) but the interaction was not significant. Tukey (b) tests indicated that eighth graders performed better than retardates ($p < .01$) and third graders ($p < .05$), who did not differ from each other. Comparisons at each redundancy level showed that

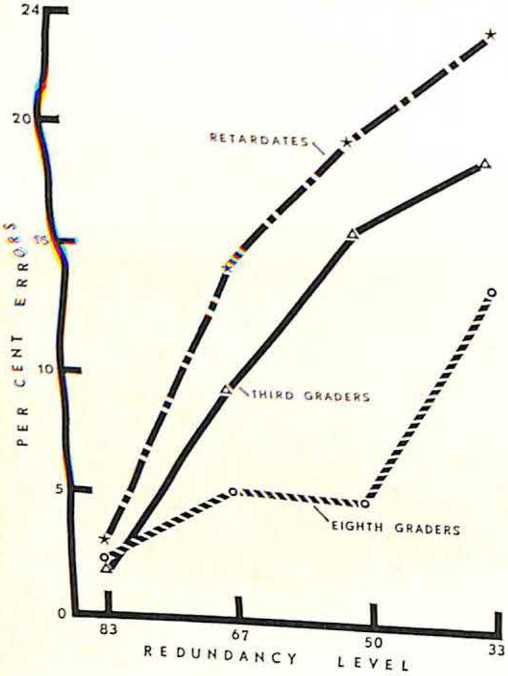


FIG. 1. Performance of three groups at four redundancy levels when paired associates were presented by the anticipation method.

eighth graders performed significantly better ($p < .05$) than retardates and third graders at the 50% redundancy level. No other comparisons between *S* groups were significant.

Since there were six paired associates in each condition, decreased errors at higher redundancy levels must result from *S*s effectively using the redundancy. How effectively the redundancy was used by each group can be deduced by comparing performance within *S* groups. For the eighth graders, the largest difference occurred between the 33% and 50% redundancy conditions ($p < .005$), with a ceiling effect limiting further improvement to a small, but significant ($p < .02$) increment in performance by *S*s on the 83% redundancy condition over *S*s in the 67% redundancy condition. For the retardates, on the other hand, the largest increment in performance occurred between the 67% and 83% redundancy conditions ($p < .02$), and there were no significant performance differences between *S*s in the 33%, 50%, and 67% redundancy conditions. The third graders were intermediate between the other groups; that is, *S*s at the 83% redundancy level performed significantly ($p < .002$) better than those at the 67% level, who in turn performed significantly better than those at the 33% level ($p < .02$). These results suggest that it took greater redundancy to make a significant impact on the retardates and—to a slightly lesser extent—on the third graders, than on the eighth graders. Put another way, in order to perform almost without error, the lower MA groups required greater redundancy than did the higher MA group.

In previous studies, 90% correct performance has been used as the criterion for successful information transmission (Spitz, 1972). The two to four units of information transmitted by retardates and normals in the present study are lower than the range usually found, and this suggests that many *S*s did not recognize the lower level redundancy, at least until later trials. This may have occurred due to difficulty in discovering the presence of redundancy because of the sequential nature of the task. If the entire set of six pairs were presented simultaneously, the redundancy might be more readily recognized. Experiment II was designed to test this possibility, and also to provide a method of revealing how effectively the redundancy was utilized.

EXPERIMENT II

The anticipation method of paired-associate learning has the presumed advantage of immediate feedback, but also has some disadvantages (Battig, 1965). On the other hand, if all six pairs were to be presented simultaneously, *S*s, while denied immediate feedback, would be free to distribute their study time in any manner they wished and, more importantly, to impose their own organization on the stimuli (e.g., Puff & Bousfield, 1967). The simultaneous method has been used infrequently, and primarily when groups of *S*s are tested (e.g., Wood, 1970).

Even less used in paired-associate learning experiments is the free-ordered recall method for testing learning of the pairs. In this method, all the stimulus terms are presented simultaneously and S is free to give the response terms in any order he wishes. Again, this method has been used in testing groups of Ss (Mayzner & Schoenberg, 1965).

The use of simultaneous presentation combined with the free-ordered recall method of testing is ideally suited to answer questions raised by Exp. I. Will Ss more readily discover the redundancy under simultaneous presentation conditions? Results from other types of learning studies suggest that they might (e.g., Puff & Bousfield, 1967). With all the pairs simultaneously presented it should be more readily apparent that—in the 67% redundancy condition, for example—two of the response terms are A's and four are B's. And with free-ordered recall, it would be most economical, once the redundancy is recognized, to respond first with the two A's and then simply respond with B's to the remaining stimuli. Of course, it is possible to recognize the redundancy and yet respond from top to bottom (or some other order), but this would not be the best strategy. Consequently, the order in which S gives his responses should reveal something about how he organizes the material.

Method

Stimuli.—The same pairs used in Exp. I were used in Exp. II, except that the 83% redundancy level condition was eliminated. In contrast to Exp. I, however, the six paired associates of each condition were presented simultaneously in a vertical array, one above the other, for the study trial blocks. The order of placement of the pairs from top to bottom duplicated the sequential order of Exp. I. A new random order was used for test trial blocks (in which stimulus terms were presented simultaneously in a vertical array).

Subjects.—There were 108 new Ss in Exp. II, 36 in each of three groups. The retardates (6 girls, 30 boys; $\overline{CA} = 15.63$; $SD = 1.56$; $\overline{IQ} = 63$, $SD = 7$) were drawn from the same population (Johnstone) as in Exp. I. The third graders (18 girls, 18 boys; $\overline{CA} = 8.55$, $SD = .27$) and eighth graders (21 girls, 15 boys; $\overline{CA} = 13.35$, $SD = .38$) were from a neighboring parochial school. No IQs were available for the normal Ss, but there is no reason to believe they differed significantly from the normals in Exp. I. The average CAs were younger than the normals in Exp. I because testing occurred earlier in the school year.

Procedure.—Pretesting was the same as in Exp. I. In the instructions, Ss were additionally told that when the numbers alone came on the screen, they could respond "in any order you like." This part of the instructions was repeated periodically throughout testing. Twelve Ss of each S group were assigned to each of the three conditions so that each subgroup was approximately equated in terms of sex and CA for the normals, and in terms of sex, CA, and IQ for the retardates. Each S then received for study six paired associates for 18 sec. (matching the 3 sec/pr of Exp. I), followed by a 7.5-sec. pause (slightly less than the average time between feedback and anticipation trial in Exp. I). This was followed by an 18-sec. block presentation of the six stimulus terms, differently randomized, with S given free recall. Since S could respond in any order he wished, it was necessary for him to say both the stimulus and the response. The 18-sec. free recall period was followed .8 sec. later by the next randomized block presentation of the six S–R pairs, and so on. As in Exp. I, the 5 trial blocks were repeated until—If S had not reached the criterion of two consecutive correct test trials—testing stopped after 30 trial blocks. Responses, and the order in which they were given, were recorded by E.

Results

As in Exp I, no sex differences were found. Results are shown graphically in Fig. 2 (where they are compared with results from Exp. I). Analysis of variance indicated that groups, conditions, and their interaction were all significant ($p < .001$). Overall, eighth graders performed significantly better than third graders, who performed significantly better than retardates ($p < .01$). Differences between retardates and eighth graders were significant at each redundancy condition ($p < .01$), between retardates and third graders were significant only at the 33% redundancy condition ($p < .05$), and between third graders and eighth graders were significant at the 67% and 33% redundancy condition (p at least $< .05$).

As in Exp. I, there were no significant differences within any S group between performance on the 50% and 67% redundancy conditions. Evidently, redundancy imposed by using two A's and four B's as response teams had no different effect on paired-associate learning than using three A's and three B's. However, for both the retardates and third graders, Ss in the 50% redundancy condition performed sig-

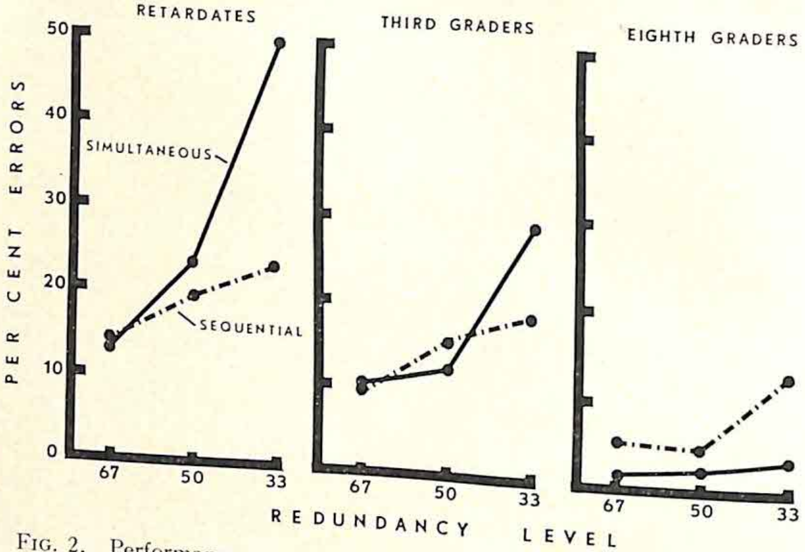


FIG. 2. Performance comparisons of three groups when paired associates were presented sequentially in Exp. I (by the anticipation method) and when they were presented simultaneously with free recall in Exp. II.

nificantly better than their peers in the 33% redundancy condition (p at least $< .05$). For the eighth graders, there were no significant differences between any of the three redundancy conditions.

Comparison with Exp. I.—In terms of overall scores, retardates performed significantly worse ($p < .005$), third graders showed no change, and eighth graders performed significantly better ($p < .001$) in Exp. II than in Exp. I. The retardates' performance decrement was due exclusively to their extremely poor performance on the 33% redundancy condition in Exp. II ($p < .005$), neither of the other two conditions differing significantly from each other. Contrariwise, the eighth graders performed significantly better ($p < .001$) on the 33% redundancy condition in Exp. II, and also on redundancy level 67% ($p < .002$), with no differences emerging on the 50% redundancy condition. The performances of the third graders were not significantly different in Exp. II than in Exp. I on any of the three conditions.

Clustering.—The following method of analyzing clustering in this task was used. One point was given each time a letter followed itself in *S*'s response order, and only when both responses were correctly paired with their appropriate stimuli. Consequently, the highest clustering score at each trial block in the 67% and 50% conditions would be 4 (A-A = 1, B-B-B-B = 3; A-A-A = 2, B-B-B = 2), and on the 33% redundancy condition would be 3 (A-A = 1; B-B = 1; C-C = 1). The percentage of clustering at each trial block was the obtained clustering divided by the total possible clustering.

It was calculated that because of the arrangement of the pairs, if *S* responded in a fixed order from top to bottom (or bottom to top) and was correct on all pairs, the average percentage of clustering would be 45, 45, and 0 for the 67%, 50%, and 33% redundancy conditions, respectively.

Figure 3 shows that on the 67% and 50% redundancy conditions, all groups made extensive use of clustering by the final two criterion trials. That is, there was a substantial increase in the number of *S*s who responded with a particular letter prior to giving responses containing the second letter. However, on the 33% redundancy condition, the picture is quite different. Only two retardates reached criterion, so retardates are not represented. The eighth graders again show a substantial increase in clustering at criterion. In fact, considering the zero base line for fixed-

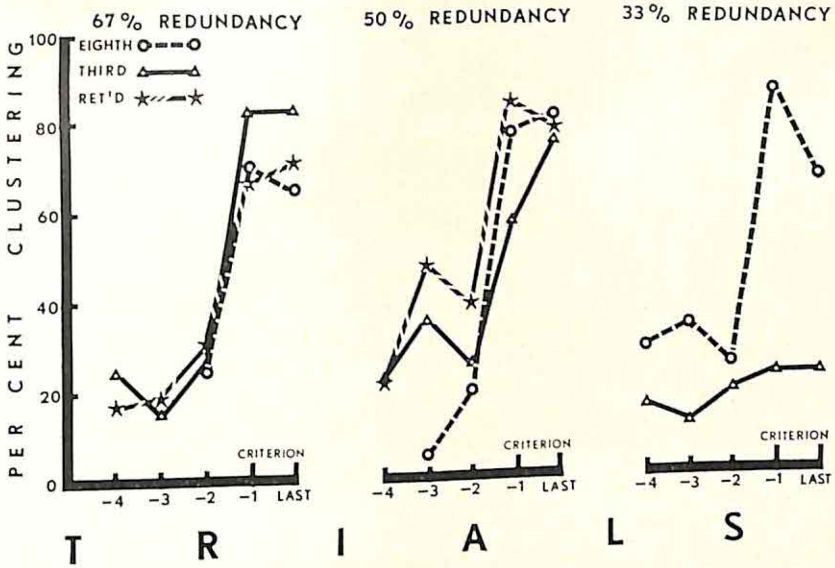


FIG. 3. Clustering of three groups in three redundancy conditions. (Since curves are plotted backwards from the point where each *S* reached criterion, only *S*s who reached criterion were used. Unplotted points had less than five *S*s (i.e., *S*s who reached criterion in less than five trials would not be represented at the −4 data point, for example). Retardates are not represented in the 33% redundancy condition because only two retardates reached criterion.)

order responding, their clustering performance is quite high. When they reached criterion, most eighth graders responded in groups of two A's, two B's, and two C's. Although third graders cluster about 20% above the zero base line, most of them reached criterion the hard way: by learning all six paired associates without making effective use of the redundancy in the material. These results strongly suggest that the major differences between *S* groups on the 33% condition is the greater capacity of the eighth graders to discover and utilize the redundancy.

DISCUSSION

The most significant finding is the rather catastrophic disintegration in performance by retardates who were presented high information material in complete form, and permitted free recall, compared to their relatively good performance when the same material was presented sequentially with immediate feedback. This contrasts strikingly with the improved performance of normals of equal CA under the same conditions. The normals of equal MA showed no change. It seems quite clear that when retardates were presented with the high information material all at once they were unable to find a successful strategy to deal with it (Spitz, 1966). The eighth graders quickly noted and made effective use of the redundancy. Of course, it is possible that the retardates' poor performance and the eighth graders' good performance were due primarily to the free-ordered recall method of testing, but it seems much more likely that the simultaneous presentation method played the major role.

The implications of these findings for the venerable whole–part learning controversy are evident: for certain kinds of material, at least, retardates will learn best when the information is presented a little at a time (as, for example, in teaching machines), while for normals of at least 13 yr. of age presentation of the same material all at once (holding time constant) allows *S*s to find and use redundancies more readily and, consequently, to perform more efficiently. This is not a totally new finding. McGeoch (1931) reviewed the relationship of mental development to whole–part learning and suggested that " 'mental development'. . . is a factor in the relative efficiency of the whole and part methods [p. 335]." The higher MA *S*s learned relatively better by the whole method than by the part method, even though the conception of whole and part methods was somewhat different from that used here. The part method was particularly good for retarded *S*s. Consonent with those findings, Madsen (1963) found that under relatively massed conditions (5-sec. intertrial interval), performance by retardates was significantly poorer than under relatively spaced conditions (60-sec. intertrial interval), an effect not found with intellectually average or above-average *S*s.

The present results supplement our series of studies on the recognition and use of redundancy by retardates. When the redundancy is relatively high, retardates can discover and utilize it (Spitz & Webreck, in press). But when the redundancy is low (i.e., 33%), they cannot and are adversely affected by simultaneous presentation of the material—so adversely affected, in fact, that they perform at a lower level than normals of equal MA.

REFERENCES

BATTIG, W. F. Procedural problems in paired-associate learning research. *Psychonomic Monograph Supplements*, 1965, **7**(1, Whole No. 1).

BAUMEISTER, A. A., & KELLAS, G. Process variables in the paired-associate learning of retardates. In N. R. Ellis (Ed.), *International review of research in mental retardation*. Vol. 5. New York: Academic Press, 1971.

BOWER, G. H. An association model for response and training variables in paired-associate learning. *Psychological Review*, 1962, **69**, 34–53.

BROWN, A. L. Subject and experimental variables in the oddity learning of normal and retarded children. *American Journal of Mental Deficiency*, 1970, **75**, 142–151.

MCGEOCH, G. O. The intelligence quotient as a factor in the whole–part problem. *Journal of Experimental Psychology*, 1931, **14**, 333–358.

MADSEN, M. C. Distribution of practice and level of intelligence. *Psychological Reports*, 1963, **13**, 39–42.

MAYZNER, M. S., & SCHOENBERG, K. M. Short-term storage and retrieval of paired-associate material. *The Journal of Psychology*, 1965, **59** 113–123.

PUFF, C. R., & BOUSFIELD, W. A. Development of intertrial organization with simultaneous presentation of stimulus items. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 213–215.

SPITZ, H. H. The role of input organization in the learning and memory of mental retardates. In N. R. Ellis (Ed.), *International review of research in mental retardation*. Vol 2. New York: Academic Press, 1966.

SPITZ, H. H. The channel capacity of educable mental retardates. In D. K. Routh (Ed.), *The experimental psychology of mental retardation*. Chicago: Aldine, 1972.

SPITZ, H. H., & BORLAND, M. D. Redundancy in line drawings of familiar objects: Effects of age and intelligence. *Cognitive Psychology*, 1971, **2**, 196–205.

SPITZ, H. H., GOETTLER, D. R., & WEBRECK, C. A. Effects of two types of redundancy on visual digit span performance of retardates and varying aged normals. *Developmental Psychology*, 1972, **6**, 92–103.

SPITZ, H. H., & WEBRECK, C. A. Effects of spontaneous vs. externally-cued learning on the permanent storage of a schema by retardates. *American Journal of Mental Deficiency*, in press.

WOOD, G. Paired-associate and free recall to free recall transfer. *Journal of Experimental Psychology*, 1970, **83**, 519–522.

(Received December 13, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 171–176

NAMING LATENCY FACILITATION: AN ANALYSIS OF THE ENCODING COMPONENT IN RECOGNITION REACTION TIME[1]

KIM KIRSNER[2]

Medical Research Council Developmental Psychology Unit, London, England

S. Sternberg's current analyses of recognition reaction time assume that encoding and memory scanning represent independent stages of information processing. This assumption may be questionable. The present study replicated the stimulus conditions used in the RRT (recognition reaction time) varied-set studies but required *S* to name the probe as rapidly as possible regardless of list membership. Naming latency facilitation (the advantage found for "old" items over "new" items) was found to increase as set size decreased both for letters and for words. This relationship was found to hold for the visual list/visual probe and auditory list/ visual probe modality combinations and at the one and two item/sec rates. It was concluded that the *encoding stage* (as distinct from the memory comparison stage) is affected by set size and that the slope of the function relating set size and RRT is therefore influenced, or possibly determined, by encoding factors.

The study of memory processes has benefited significantly from the development of a technique (Sternberg, 1967, 1969a, 1969b) which provides independent estimates of the duration of *encoding* and *memory comparison* operations in recognition memory. The technique has been successfully applied to a variety of issues including the nature of coding in short-term memory (Chase & Calfee, 1969; Sternberg, 1967), developmental changes (Hoving, Morin, & Konick, 1970), and the lateralization of memory for verbal and pictorial stimuli (Klatzky & Atkinson, 1970). The present investigation was implemented in order to assess some of the assumptions underlying the distinction between encoding and memory comparison operations in recognition memory.

The memory comparison component of RRT (recognition reaction time) is estimated from the slope of the function relating RRT and memory set size. The encoding component is included in the zero intercept of this memory or "scanning" function. However, the validity of the foregoing analysis rests on an assumption that the set-size variable influences only the memory comparison stage of the recognition process. That is, if set size has an effect on the encoding stage of the recognition process, it follows that the intercept and slope values do not provide independent estimates of the encoding and memory comparison stages, respectively. In a previous study, Kirsner and Craik (1971) found that the serial position of the target item in a list of words gave rise to systematic changes in the time taken to identify the item when it was repeated as a probe.

Specifically, naming latency for probe items was shorter when they were drawn from the last two to three serial positions in the list as compared with "new" items. (In their experiment, *S*s named the probe item regardless of whether it had occurred in the list or not.) On the basis of these results, it seems likely that naming latency will be sensitive to set size and, therefore, that a component of the scanning function in RRT may be due to encoding as distinct from memory factors. The present study examines directly the relationship between set size and naming latency facilitation.

A further question concerns the role of sensory factors in the RRT paradigm. On the basis of their findings that the scanning function was flatter when the list and probe items were presented in the same modality

---

[1] This research was supported by Medical Research Council Grant G77/3097.

[2] Requests for reprints should be sent to Kim Kirsner, Department of Psychology, University of Toronto, Toronto 5, Ontario, Canada.

(e.g., visual list/visual probe or VV) as compared with different modalities (e.g., auditory list/visual probe or AV), Chase and Calfee (1969) concluded that the recognition comparison may be based on information retained at a sensory level. They used letters as stimuli. However, Kirsner and Craik (1971), using eight-item lists of words, found no differences at any serial position between the VV and AV combinations (in either naming latency facilitation or RRT). This discrepancy may reasonably be attributed to the different stimuli used in the two experiments. That is, visual sensory memory may be restricted in capacity to two to three letters and VV would not therefore facilitate RRT or naming latency when the stimuli used were words as in the study by Kirsner and Craik. A direct test of this question is made in the present study by including a comparison between naming latency facilitation for letters and words using both the VV and AV modality combinations. If naming latency facilitation is determined at a sensory level and the relevant store can hold information about two to three letters only, then this should be apparent as an advantage for VV in the letter condition.

Another question concerns the role of word-identification processes in short-term memory. Although a distinction is commonly drawn between prelinguistic and linquistic memory, little detailed consideration has been given by memory theorists to the role of the interpolated identification mechanism in their experiments. It seems plausible, however, that units in the word-identification system may themselves retain persisting evidence of recent activation, and that this information could contribute to subsequent item recognition. Morton (1970) has described some of the characteristics of the units, or logogens, in the word store. On the basis of evidence from word perception and short-term memory experiments, Morton suggested (*a*) that logogen "priming" may occur irrespective of the source of stimulation, and (*b*) that the magnitude of priming is reduced by subsequent stimulus events. Accordingly, if naming latency facilitation reflects logogen priming, it follows that facilitation should be equal for VV and AV, and that it should show systematic serial position differences including a recency component.

Although the stimulus conditions pertaining to naming latency and RRT may be identical, this does not guarantee that the processes governing naming latency facilitation also affect RRT. Specifically, the pattern of naming latency facilitation could be determined by response preparation factors which cannot operate in the RRT situation, where there are only two possible responses at each set size. In an attempt to obtain some information on this question, rate of presentation was included as an experimental variable. If response preparation is critical in naming latency facilitation, then, contrary to the sensory prediction, facilitation should be greatest at the slow presentation rate.

To summarize, the experimental design comprised the factorial combination of stimulus material (letters vs. words), list presentation modality (VV vs. AV), presentation rate (2 items/sec vs. 1 item/sec), and set size (one to four items). If naming latency facilitation is influenced by sensory factors, facilitation should be greatest for VV letters at the fast presentation rate. However, if response preparation is critical, facilitation should be insensitive to list presentation modality, and greatest at the slow presentation rate.

METHOD

Design.—The experimental design was a 2 × 4 × 2 × 2 × 4 crossed classification in which the variables were stimulus material, *S*s, list presentation modality, presentation rate, and set size. Thus stimulus material was a between-*S* variable.

Each of the eight *S*s completed four sessions, one session for each combination of list presentation modality and presentation rate. Each session comprised four 52-trial blocks, one block per set size. List presentation modality and presentation rate orders were counterbalanced across *S*s, and set-size order was counterbalanced across the other experimental conditions and *S*s.

The first 4 trials were designed to absorb short-term practice effects and were not included in the analysis. There were 48 experimental trials. The probe item was "old" (from the immediately preceding list) on 24 of these trials. The old trials were divided equally, and randomly, among the serial positions pertaining to the set size for the block of trials. Each block used a different random sequence of old trials and there were two formats for each combination of the experimental conditions.

Procedure.—Each *S* was tested individually in four sessions. The instructions were: "Attend to each word in the list as it is presented, but do not repeat the words to yourself or rehearse them again before the test word is presented. Concentrate on naming the test word as rapidly as possible regardless of whether it was in the list or not."

The *S* sat in a sound-reduced booth. Auditory list items, previously recorded in a female voice, were presented through headphones from one track of stereo tape recorder (Uher Royal De Luxe). The visual list and probe stimuli (in uppercase print) were inserted in an IBM typewriter, which was advanced by a solenoid attached to the return button. The solenoid was activated by a digitimer. In the AV condition, the probe cycle was coordinated by a diapilot signal on the tape. A Shibaden TV camera was focused on the typewriter carriage, and *Ss* viewed the visual stimuli on a Shibadan monitor. Latencies were obtained using a throat microphone.

In the slow presentation rate, each list stimulus appeared for .5 sec. and was followed by an .5 sec. blank interval before the appearance of the next list item. The offset/onset probe delay interval was 1.5 sec. These conditions were exactly halved at the fast presentation rate.

Material.—The following letters were used: B, D, F, G, J, K, L, M, N, P, T, V, and Z. The words were randomly selected from a pool of 300 items. The word pool for the present experiment was restricted to two-syllable nouns of five, six, or seven letters with homophones, proper nouns, contractions, and archiac words deleted. Only words beginning with the letters listed above were included. The word lists were constructed by sampling randomly without replacement within blocks but with replacement across blocks. Letter lists were constructed by sampling randomly with two constraints: (*a*) that each letter should be used as a probe on three, four, or five trials within a block, and (*b*) that no letter should occur on two occasions in one trial list.

Subjects.—The *Ss* were obtained from the Birkbeck College pool. All *Ss* were between 18 and 31 yr. of age.

RESULTS

The magnitude of naming latency facilitation was examined in a series of five-way ANOVAs (Stimulus Material × *S* × List Presentation Modality × Presentation Rate × Set Size or Serial Position). Following the analysis used by Kirsner and Craik (1971), the value used in each ANOVA was mean naming latency for new items minus mean naming latency for old items from each condition (that is, the difference resulting from the occurrence of an item in the immediately preceding list). This technique provides a control over the nonspecific effects of factors such as practice and attention on RT. Electrical and mechanical failures accounted for approximately 1% of the responses.

An ANOVA showed that the effects of set size, $F\ (3,\ 9) = 54.77$, $p < .01$, and stimulus material, $F\ (1,\ 3) = 10.18$, $p < .05$, were significant, but that no other effects approached statistical significance. The results, pooled across presentation, modality, and rate, are shown in Fig. 1, where it is apparent that the magnitude of naming latency facilitation increases as a function of decreasing set size, and naming latency facilitation is greater for words than letters. A significant linear trend underlay the set-size function, $F\ (1,\ 9) = 134.6$; $p < .01$), and further analysis showed that the quadratic component was also significant, $F\ (1, 9) = 28.6$; $p < .01$. The linear and quadratic components accounted for 81.7 and 17.4% of the variance, respectively.

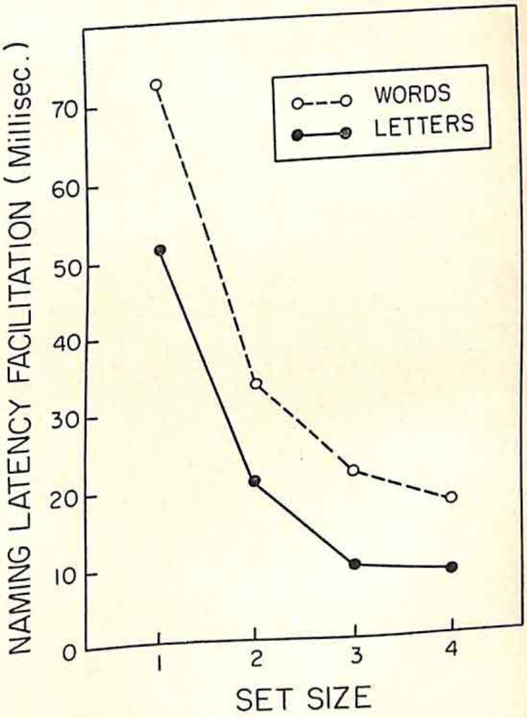


FIG. 1. Naming latency facilitation as a function of set size and stimulus material.

TABLE 1

Naming Latency Facilitation as a Function of Set Size

Set size	Material	Backward serial position			
		4	3	2	1
4	letters	5.3	2.2	4.5	22.4
	words	13.1	10.9	14.1	27.9
3	letters		3.0	15.5	14.0
	words		13.8	19.7	30.2
2	letters			21.2	17.8
	words			30.5	36.0
1	letters				51.5
	words				72.4

Table 1 shows the naming latency facilitation data classified by stimulus material, set size, and serial position. It is apparent that the set-size effect is not an artifact of differences in relative recency between groups. A second ANOVA confirmed this conclusion by showing that the effect of set size was statistically significant when based solely on data from the last serial position in each set, F (3, 9) = 34.8; $p < .01$. A third ANOVA using data from Set Size 4 only showed that the serial position differences (for Set Size 4) presented in Table 1 are statistically reliable, F (3, 9) = 8.6; $p < .01$).

The naming latency values for new items were examined for any possible artifacts due to set size, modality, rate, or practice. The relevant latencies were 573, 578, 574, and 571 for Set Sizes 1, 2, 3, and 4; 569 and 579 for VV and AV; 572 and 576 for the fast and slow rates; and 579, 573, 578, and 566 msec. for Sessions 1, 2, 3, and 4, respectively. For each comparison, the data were pooled across all other variables. The latencies for new letters and new words were 546 and 601 msec., respectively.

Discussion

The pattern of the results was straightforward. The magnitude of naming latency facilitation increased as a function of decreasing set size for both letters and words. This effect was insensitive to list presentation modality and presentation rate, but overall facilitation was greater for words than letters. The slope of the function relating set size and naming latency facilitation was 15.8 msec. per item, however this value is qualified by the presence of a significant quadratic component apparently due to an additional increase in facilitation at Set Size 1.

In general terms, naming latency facilitation may be attributed to "sensory," "identification," or "response" factors. According to the sensory hypothesis (Eichelman, 1970; Parks, Kroll, Salzberg, & Parkinson, 1972; Posner, 1971), facilitation of RRT or naming latency should be at a maximum when the list and probe stimuli are physically as opposed to nominally identical. The conditions for this comparison are met in the present experiment where VV and AV can be said to represent physical and nominal identicality, respectively. The results are clearly inconsistent with the sensory hypothesis because no advantage was found for VV over AV at the fast or slow presentation rate for either letters or words.

According to one possible response preparation position, naming latency facilitation may be attributed to articulatory priming. That is, naming latency is facilitated when *S* articulates the target item immediately prior to presentation of the probe. The presentation rate manipulation was included in order to test this hypothesis. Specifically, it was thought that when two or more words are presented at the fast presentation rate, *S*s would have little time for articulatory preparation and, therefore, that naming latency facilitation would be attenuated. The results did not support this response preparation position.

In the identification account of naming latency facilitation, it is assumed that the internal representation or logogen (Morton, 1970) of the letter or word may be transiently primed by recent experience and thus more easily "fired" on a second, test presentation. This notion receives some support from the advantage found for words over letters. As each letter is presented as an untested list item or as a probe item between 7 and 17 times per block in the letter condition, it seems likely that the threshold level of all sampled letters will be lowered accordingly. Thus there will be less opportunity for additional transient facilitation. The modality and serial position results are also consistent with the logogen model which specifies that priming can occur irrespective of the source of stimulation and that the amount of activation in the logogen will be reduced by subsequent stimuli.

The additional facilitation found in the Set Size 1 condition may be analagous to that reported for positive items in a number of RRT studies (Burrows & Okada, 1971; Clifton & Birenbaum, 1970; Forrin & Morin, 1969). On the input side, one possible explanation is that the effect is due to residual sensory facilitation in VV whereas, in AV, *S*s generate a visual matching code from the auditory signal. The possibility of this transformation has been shown for letters by Posner, Boies, Eichelman, and Taylor (1969). However, it does not seem likely that *S*s could generate a visual matching code for a five- to seven-letter word in .75 sec. Weber and Bach (1969), for example, estimated the visual synthesis rate at two letters per sec. The synthesis hypothesis is not entirely precluded by the present data however. On the output side, it is possible that the advantage arises in response organization. That is, in the Set Size 1 condition, *S* is prepared to emit the list item when the probe stimulus is presented. If this view is correct, it might be expected that naming latency for new items would be increased in the Set Size 1 condition. This does not occur. Thus the evidence may be consistent with a word store interpretation of the Set Size 1 effect also. That is, the relevant logogen is primed, possibly as a function of the amount and, to a lesser extent, the recency of attention, and the advantage therefore occurs in the final stage of item identification. At this level, however, there may be no valid distinction between perceptual and response processes. Thus item identification may be functionally equivalent to the act of response selection.

The presence of a relationship between set size and naming latency appears to pose some problems for those models of recognition memory which assume (*a*) that encoding and scanning represent independent information-processing stages in recognition memory, and (*b*) that the intercept and slope values of a function relating set size and RRT provide independent estimates of these stages. If, as has been argued above, naming latency facilitation occurs during item identification, then it follows that a component (not more than 16 msec. per item) of the function relating set size and RRT may be attributed to a change in the duration of the encoding stage prior to memory analysis. That is, the slope value in RRT is a compound of encoding and memory comparison factors. A second possibility is that changes in both naming latency and RRT enjoy common determination in the logogen system. In this view the activity level of the logogen of a word presented in a given set is determined by the serial position of that item in the set, and also by the number of items in the set. Naming latency is then a direct "index" of the activity level of the logogen of the test word. In recognition memory, on the other hand, the logogen's activity level acts as the input into a decision mechanism where it is compared with some criterion. Thus, as in the trace strength account of recognition memory (Baddeley & Ecob, 1970; Norman & Wickelgren, 1969), RRT will reflect the amount by which the input exceeds or, in the case of negative probe items, falls short of the criterion.

Conclusions.—The present study was implemented to consider the possibility that in Sternberg's (1967) RRT paradigm the set-size variable may have an effect on encoding as well as subsequent memory comparison operations. As in the RRT paradigm, *S*s viewed lists of items which were followed immediately by a probe item. The *S*s named the probe item as rapidly as possible regardless of list membership. The results showed that naming latency, for both letters and words, is sensitive to the number of items in the preceding list and also to the serial location of the probe item in the list. The magnitude of naming latency facilitation was insensitive to list presentation modality and presentation rate. The results were thought to be consistent with the view that set size does have an effect on encoding under stimulus conditions similar to those used in Sternberg's (1967) character classification studies, however further research is required to determine whether these changes in encoding determine, contribute to, or are irrelevant to the set-size effect found for RRT.

REFERENCES

Baddeley, A. D., & Ecob, J. R. *Reaction time and short-term memory: A trace strength alternative to the high-speed exhaustive scanning hypothesis.* (Tech. Rep. No. 13) La Jolla, Calif.: University of California, Center for Human Information Processing, 1970.

Burrows, D., & Okada, R. Serial position effects in high-speed memory search. *Perception & Psychophysics*, 1971, **10**, 305–308.

Chase, W. G., & Calfee, R. C. Modality and similarity effects in short-term recognition memory. *Journal of Experimental Psychology*, 1969, **81**, 510–514.

Clifton, C., & Birenbaum, S. Effects of serial position and delay of probe in a memory scan task. *Journal of Experimental Psychology*, 1970, **86**, 69–76.

Eichelman, W. H. Stimulus and response repetition effects for naming letters at two response–stimulus intervals. *Perception & Psychophysics*, 1970, **7**, 94–96.

Forrin, B., & Morin, R. E. Recognition times for items in short- and long-term memory. *Acta Psychologica*, 1969, **30**, 126–141.

Hoving, K. L., Morin, R. E., & Konick, D. S. Recognition reaction time and size of the memory set: A developmental study. *Psychonomic Science*, 1970, **21**, 247–248.

Kirsner, K., & Craik, F. I. M. Naming and decision processes in short-term recognition memory. *Journal of Experimental Psychology*, 1971, **88**, 149–157.

Klatzky, R. L., & Atkinson, R. C. Memory scans based on alternative test stimulus representations. *Perception & Psychophysics*, 1970, **8**, 113–117.

Morton, J. A functional model for memory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

Norman, D. A., & Wickelgren, W. A. Strength theory of decision and latency in retrieval from short-term memory. *Journal of Mathematical Psychology*, 1969, **6**, 192–208.

Parks, T. E., Kroll, N. E. A., Salzberg, P. M., & Parkinson, S. R. The persistence of visual memory as indicated by decision time in a matching task. *Journal of Experimental Psychology*, 1972, **92**, 437–439.

Posner, M. I. Abstraction and the process of recognition. In T. G. R. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation*. New York: Academic Press, 1971.

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology*, 1969, **79** (1, Pt. 2).

Sterngerg, S. Two operations in character recognition: Some evidence from reaction-time experiments. *Perception & Psychophysics*, 1967, **12**, 45–63.

Sternberg, S. The discovery of processing stages: Extensions of Donder's method. *Acta Psychologica*, 1969, **30**, 276–315. (a)

Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, **57**, 421–457. (b)

Weber, R. J., & Bach, M. Visual and speech imagery. *British Journal of Psychology*, 1969, **60**, 199–202.

(Received December 15, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 177–183

TIME COURSE OF IDENTITY AND CATEGORY MATCHING BY SPATIAL ORIENTATION[1]

MERRILL F. ELIAS[2]

Center for the Study of Aging and Human Development and Department of Psychiatry Duke University Medical Center

MARCEL KINSBOURNE

Division of Pediatric Neurology, Duke University Medical Center

The *S*s responded SAME or DIFFERENT to a form varying only in spatial orientation. Onset of one orientation (probe) followed another (target) by 100, 600, or 1,200 msec. For identity matching, SAME responses were given to target-probe combinations of identical orientation. For category matching, SAME responses were given to target-probe combinations falling in a common set. Category matching was less rapid than identity matching, but the time course of matching differed from that obtained previously with verbal materials. The former became more efficient. The latter remained constant. Eliminating the necessity for generating the alternate member of the set to match with the probe did not enhance matching efficiency. Findings were related to serial and parallel processing operations imposed by the stimuli.

In even the most elementary task of categorizing a stimulus as a member of a set of stimuli, *S*s must keep alternative criteria for classification in mind, and these alternatives compete for attention. Posner and Boies (1971) have summarized a number of studies which have shown that a sameness–difference judgement for two stimuli presented in temporal succession varies in latency depending on whether it is done as a physical match or a match in terms of a category the two stimuli share in common. This variation appears to be related to the differences in time required to preprocess the first stimulus in anticipation of the second. The present investigation explores the manner in which capacity for preprocessing is shared between alternative matching tasks defined by the spatial relationships of the stimuli to be matched.

Matching operations have been extensively investigated with stimuli that have a high potential for verbal and auditory encoding (Boies, 1969; Dainoff, 1970; Posner, 1969; Posner, Boies, Eichelman, & Taylor, 1969; Posner & Keele, 1967; Posner & Taylor, 1969). Multiple stimuli, e.g., A, a B, b, are presented in temporal succession, and *S*s view one as target and one as probe. Under identity-matching instructions, they respond SAME when the target and probe are identical and DIFFERENT when they are nonidentical. Under category-matching instructions, they respond SAME when the target and probe have the same name, e.g., A—a, and DIFFERENT when they do not. In these studies, reaction time to the probe has been related to the nature of the match, instructions, and length of the target-probe interval. When the target disappears during the target-probe interval (Posner & Boies, 1971), matching latencies are greater under category-matching instructions than identity-matching instructions; there is also a rapid increase in time required for physical matching with increasing target-probe interval, while category-matching latencies remain relatively constant. Posner and Boies (1971) argue that the difference in time for physical and category matching is related to the inability of *S*s to handle certain mental tasks simultaneously even when they wish to do so (limited central processing capacity). Further, they provide evidence that the slowing of physi-

---

[1] This research was supported in part by Research Training Grant 5 TO1 HD00164 from the National Institute of Child Health and Human Development. Many thanks are extended to Lucinda A. Eitreim, Division of Pediatric Neurology, for her very valuable assistance in the running of *S*s and the analysis of data. Appreciation is extended to D. Ramm, C. F. Starmer, and G. R. Marsh, Duke University Medical Center, for their suggestions and assitance with regard to data processing by computer.

[2] Requests for reprints should be sent to Merrill F. Elias, Department of Psychology, West Virginia University, Morgantown, West Virginia, 26506.

cal matching with increasing target-probe interval is due to difficulty in maintaining a viable visual image of the target, and that this difficulty is, in turn, the result of competition for attention by physical- and category-matching tasks. However, the time course of physical and category matching is related to conditions which bias *S*s' attention to the name code of the letters and away from their physical characteristics. When *S*s were provided with increased incentive to attend to the visual aspects of the target-probe stimuli by presentation of "pure" target-probe combinations, i.e., only uppercase letters (Posner, 1969; Posner & Boies, 1971), the slowing of physical-identity matching over time was more gradual, and the time difference between identity and category matching was largest at the shortest target-probe interval.

Although Posner (1969) was able to increase attention bias toward the physical characteristics of letters, attention was still strongly biased in the direction of their name characteristics as evidenced by the fact that category (name)-matching efficiency changed very little over time, and pure physical matches continued to exhibit a decrement with increasing target-probe interval. This is to be expected. Verbal materials, by nature of their symbolic function in spoken and written communication, readily lend themselves to categorization by name and to encoding in the acoustic modality (Dainoff, 1970). It seems, then, that an important test of the Posner and Boies (1971) "attention model" is a comparison of matching efficiency for stimuli which lend themselves much less readily to name and/or acoustic encoding than upper- and lowercase letters. This was accomplished in the present experiments by presenting physically identical target-probe materials which differed only in spatial orientation and by requiring *S*s to categorize them in a manner that was cumbersome, if not impossible, to express in words or to encode acoustically. Based on the Posner and Boies attention model, the following hypotheses were advanced. Category- and identity-matching tasks would compete for attention, with the former requiring more time than the latter, particularly at the shortest target-probe interval. Due to the considerably greater attentional bias toward the representational aspects of the stimuli, and in contrast to findings for verbal materials (Posner & Boies, 1971), matching for identical target-probe combinations would not decrease in efficiency with increasing target-probe interval, regardless of whether *S* is matching under identity or category instructions. Identity matching would take temporal priority over category matching in the sequence of preprocessing events prior to arrival of the probe. With increasing target-probe interval, category matching would become more rapid and approach identity matching efficiency due to the increased time available to attend to category as well as physical information prior to arrival of the probe.

EXPERIMENT I

Method

Subjects, apparatus, and procedure.—Ten females and 2 males, 19 to 32 yr. old, served as *S*s. They were paid volunteers from a population of Duke University students and medical technicians.

Stimuli were presented in a Scientific Prototype (Model GB) three-channel tachistoscope. The stimulus population is shown in Fig. 1. One of these four orientations of the same form was presented on each trial as the target and one as the probe. The interstimulus interval (ISI) was 100, 600, or 1,200

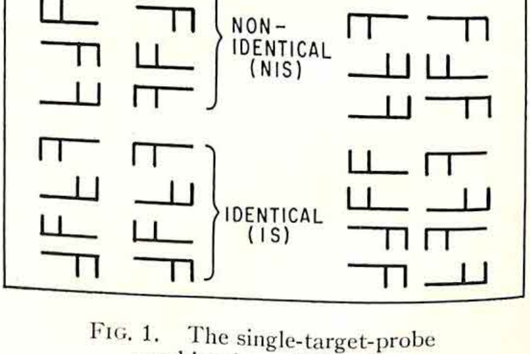


FIG. 1. The single-target-probe combinations of Exp. I.

msec. Target and probe exposures were 50 and 100 msec., respectively. White 12.7 × 17.8 cm. notecards, illuminated with 25 ftl., served as backgrounds for the stimuli, the blank ISI interval, and the fixation point viewed prior to each sequence of target, ISI, and probe. Targets were centered on their backgrounds with the center of the probe appearing 9′ (visual angle) to the right and 9′ below the center of the target. The forms were black stick figures with prong and stem lengths of approximately 21′ and 34′.

The *S*s depressed microswitches with their preferred and nonpreferred hands prior to each trial. Six *S*s lifted the index finger of the preferred hand for SAME responses and the index finger of the nonpreferred hand for DIFFERENT responses. Six *S*s followed the opposite procedure. A Standard Electric timer began with the onset of the probe and terminated with *S*'s correct response. Each *S* performed under identity- and rule-matching instructions. Identity instructions required a SAME response for the four target-probe combinations of identical orientation (Fig. 1), and a DIFFERENT response for nonidentical orientations. Category matching required a SAME response if the target and probe were common members of one of the two sets shown in Fig. 1 and a DIFFERENT response if they were members of different sets. The *S*s were shown the sets prior to testing and reminded that target-probe combinations of identical orientation were also in the same set. Each trial began with a verbal ready signal from .5 to 1 sec. before target onset and ended with a response to the probe. Time between trials was approximately 15 sec. Feedback concerning errors was provided, and *S*s were strongly encouraged not to sacrifice speed for the sake of accuracy, but to make SAME and DIFFERENCE judgments as quickly as possible.

Design.—All *S*s performed under category- and identity-matching instructions on alternate days. Six days of identity matching and 6 days of category matching were given with half of the *S*s starting with identity matching and half with category matching. Each of the ISIs was presented at each session and counterbalanced so that each *S* received every possible ordering of ISI for category- and identity-matching sessions, respectively. To equate the proportion of SAME and DIFFERENT responses for the identity-matching condition, the four identical target-probe combinations were given three times under each ISI for each session, and the 12 nonidentical combinations were given once. For category matching, the 8 combinations requiring DIFFERENT responses were given twice for each ISI. Target-probe combinations appeared in a randomly determined sequence at each ISI.

Three conditions of category matching and two conditions of identity matching, defined by the target-probe combinations (Fig. 1), were of interest: target and probe identical and in the same set, category matching (IS); target and probe nonidentical and in the same set, category matching (NIS); target and probe nonidentical and in different sets, category matching (DS); target and probe identical, identity matching (I); target and probe nonidentical, identity matching (D). Thus, a 5 (Target-Probe Conditions) × 3 (ISI) completely-repeated-measurements design was employed.

Results

Reaction time (RT) reached an asymptotic level and error rates were negligible after the second and third sessions for the identity- and category-matching tasks, respectively. Data are reported for the remaining sessions with 288 SAME and 288 DIFFERENT responses for each *S* for both tasks. Analyses of variance were done with a completely-repeated-measurements model utilizing *S*s × Effect interactions as denominators for each *F* ratio. There was a significant main effect for target-probe conditions, $F(4, 44) = 6.59, p < .01$. To avoid the necessity of assuming symmetry of variance–covariance matrices (Winer, 1962, pp. 123, 313, 345, 636), specific contrasts were done with a multivariate statistical program (Starmer, 1967). All univariate tests were done with degrees of freedom of 1 and 11. The *F* values at $p < .05$ and $p < .01$ were 4.84 and 9.65, respectively; however, following Cole and Grizzle's (1966) procedure, univariate *F*s were not reported as significant unless the multivariate *F* approximation of Hotelling's T^2 was significant, $p < .05$, for all contrasts. Contrasts between target-

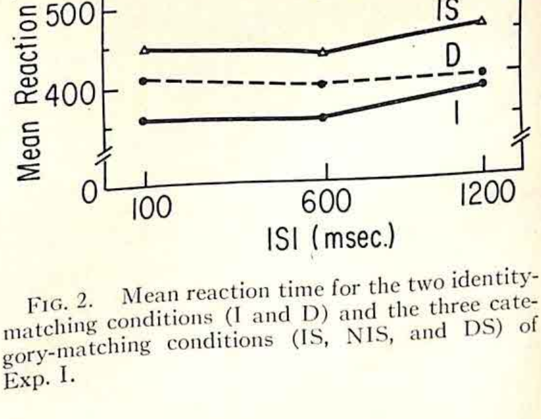


FIG. 2. Mean reaction time for the two identity-matching conditions (I and D) and the three category-matching conditions (IS, NIS, and DS) of Exp. I.

probe conditions, averaging scores for ISI, indicated significant differences between the following pairs of means: I-D, $p < .05$; I-IS, $p < .05$, I-NIS, $p < .01$; D-NIS, $p < .01$; NIS-DS, $p < .01$.

Neither the ISI main effect, $F(2, 22) = 2.21$, nor the ISI × Target-Probe Condition interaction, $F(8, 88) = 1.91$, both $p\text{s} > .05$, were significant. However, specific a priori contrasts (Winer, 1962, p. 208) were done for each target-probe condition separately since the time course of identity and category matching was the central experimental question. Observed differences between means were nonsignificant, $p > .05$, for all conditions but the NIS condition. Here, RT was significantly longer for the 100-msec. ISI than for the 600- and 1,200-msec. ISI, both $p\text{s} < .05$, but the latter two means did not differ significantly, $p > .05$.

Discussion

Posner and Keele (1970) argue that certain classes of mental operations, i.e., response initiation, rehearsal, and mental transformation, cannot be stimultaneously accomplished with equal efficiency within the same "time space." It appears that these mental operations are involved in category and identity matching by spatial orientation as well as by letter name and identity (Boies, 1969; Dainoff, 1970; Posner, 1969; Posner & Keele, 1967; Posner & Taylor, 1969). Consistent with findings for the latter materials, identity and category matching by spatial orientation did not begin and end at the same time, although the two operations may have overlapped in time. Identity matching appears to have received a temporal priority in the sequence of events which preceded arrival of the probe. When time to prepare for the probe was most severely limited, category-matching latencies were most clearly prolonged. Even under instructions to match by category, matching was more efficient when the target and probe were identical. Posner (1969) and Morton (1969) suggest that this phenomenon reflects an activation of internal units representing the stored information which become relatively more sensitive to identical stimulation for some period of time. Instructions to match by category did, however, distract attention from the physical characteristics of the target, as evidenced by shorter latency of response for identical target-probe combinations presented under identity- as opposed to category-matching instructions. This finding for spatial materials lends further support to Posner and Boies' (1971) argument that maintenance of a viable visual image of the target is related to the extent to which attention is directed toward the physical characteristics of the target during the retention interval.

As hypothesized, the time course of identity and category matching did differ from that which has been reported for verbal materials. For letter matching, significant increase in time required to make physical matches accompanied increasing ISI when the target was not present during the ISI (Posner & Boies, 1971). These investigators argue that this phenomenon reflects the possibility that *S*s keep the letter name (category) in mind during the ISI. In the present experiment, in contrast to findings for letter matching, identity matching remained constant over time while one condition of category matching, NIS, became more efficient with increasing ISI. It appears, then, that visual information is accessible, or viable, for longer periods of time for spatial materials which do not afford the opportunity for acoustic encoding (Dainoff, 1970) or categorization by name (Posner, 1969; Posner & Boies, 1971).

EXPERIMENT II

With increased time to prepare for the probe, category-matching efficiency increased under the NIS condition. This phenomenon suggests that a mental operation, in preparation for the probe, goes on during the ISI and is more complete after 600 msec. than it is within 100 msec. In fact, it may be as complete as possible after 600 msec. since additional time did not facilitate matching further. Experiment II was designed to (*a*) replicate this finding and (*b*) determine the nature of this preprobe activity. Increase in category-matching efficiency has been reported for verbal materials but only when the target letter remained visable during the ISI (Boies, 1970). Boies (1970), Posner and Keele (1970), and Posner and Boies (1971) suggest that the increase in category-matching efficiency with increasing ISI is due to increased time to generate visual descriptor of the opposite case, i.e., A—a, during the target-probe interval. By analogy, it is possible that *S*s in the present experiment utilized the preprobe interval to generate an alternate member of the positive target set, possibly by mental rotation of the target stimulus (Shepard & Metzler, 1971). To test this "generation" hypothesis, both the target and its alternate member were presented as a target (dual target) and compared with the single-target condition. The logic of this comparison was as follows: If generation of the alternate member of the set is a time-consuming process which slows category matching at 100 msec., removing the necessity for this operation, by presenting the entire set as target, should enhance matching efficiency at 100 msec. Thus, matching with dual targets should be more efficient than matching with single targets at 100 msec., and the increase in matching efficiency with increasing ISI should not be observed.

Method

Twelve unpracticed *S*s of the same age and sex distribution as in Exp. I were used. As may be seen in Fig. 3, a dual-target condition was created by adding the missing member of the positive set to each of the single-target combinations. To eliminate the possibility of masking and/or confusion resulting from the overlap of three stimuli, the dual-target members were separated by 17′ visual angle (end to end), and the probe was separated from the right member of the dual target set by 17′. The fixation point was located midway between the dual-target members. For the single-target condition, the target was located in the same position as the left member of the dual-target set and the probe was located in the same position as the probe for the dual-target condition. Dual and single targets were presented in alternate, counterbalanced succession for six practice and six test sessions.

Design and procedure.—For each ISI there were the three single-target-probe categories of Exp. I, IS, NIS, and DS; and three dual-target categories, Dual-IR (right member of target and probe identical), Dual-IL (left member of target and probe identical), and Dual-DS (neither member of the target identical to the probe). Exactly the same procedure as in Exp. I was employed except that *S*s were told to match by category with no mention of identity matching.

Results

Figure 4 shows mean RT under each experimental condition. Mean-effect ISI, main-effect target-probe condition, and ISI × Target-Probe Condition interaction were all significant, $F\ (2,\ 22) = 6.48$, $p < .01$; $F\ (5,\ 55) = 2.47$, $p < .05$; $F\ (10,\ 110) = 2.40$, $p < .05$. Specific contrasts with Hotelling's T^2 test indicated significantly shorter RT at the 100-, 600-, and 1,200-msec. ISIs, $p\text{s} < .05$, for the IS as opposed to all other conditions. The Dual-DS condition resulted in significantly longer RT than the Single-DS condition and Dual-IL condition for the 100-msec. ISI, $p\text{s} < .05$. The Dual-IR and Dual-IL conditions did not differ significantly for any

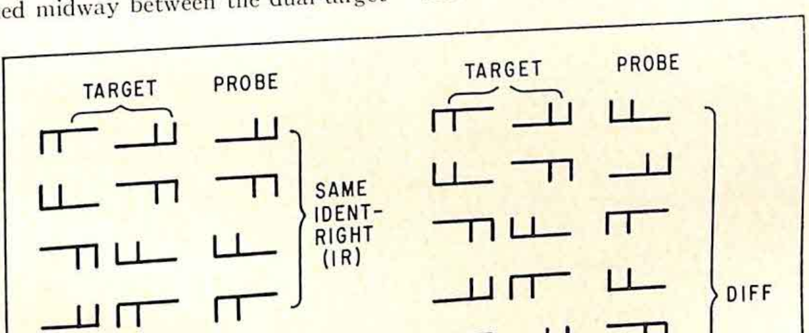


FIG. 3. The dual-target-probe combinations of Exp. II.

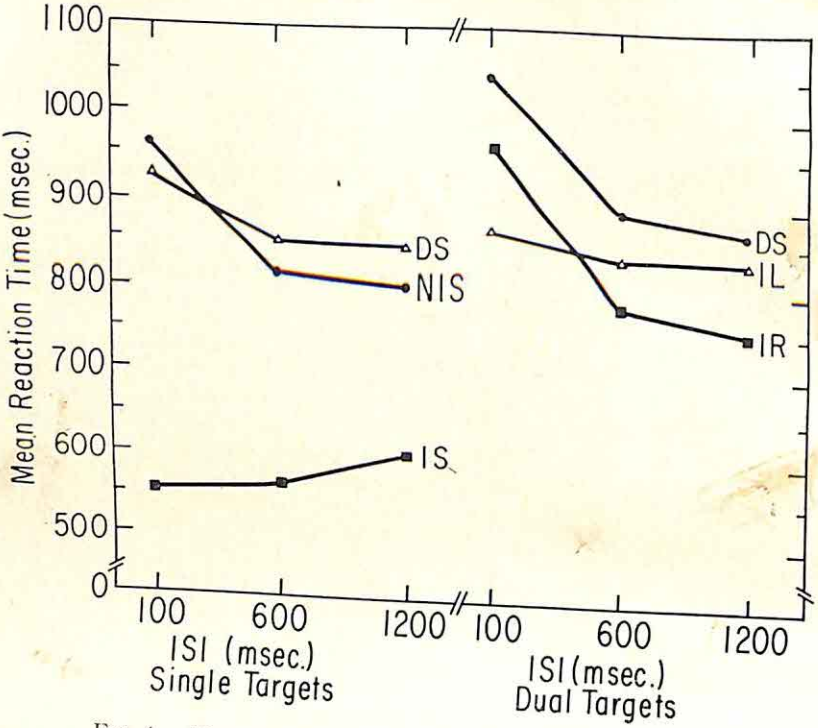


FIG. 4. Mean reaction time for the single- and dual-target categories of Exp. II.

ISI, $p\mathrm{s} > .05$. Contrasts between ISI means for each target-probe condition indicated significantly shorter RT for the 600- and 1,200-msec. ISI than for the 1,200-msec. ISI for the Single-NIS condition, $p\mathrm{s} < .05$, the Dual-DS condition, $p\mathrm{s} < .05$, the Single-DS condition, $p\mathrm{s} < .05$, and the Dual-IR condition, $p\mathrm{s} < .01$. All other contrasts revealed significant differences, $p\mathrm{s} > .05$.

Discussion

Experiment II does not support the hypothesis of impaired matching efficiency at the shortest ISI because of insufficient time to generate the alternate member of the set prior to arrival of the probe. For the 100-msec. ISI, "difference" RT was actually longer for dual as opposed to single targets. For each ISI, matching was slower for Dual-IL and Dual-IR conditions as opposed to the Single-IS condition. Two possibilities arise for verification through further experimentation: (*a*) the hypothetical process of "generation" is accomplished within 100 msec. for materials varying in spatial orientation; (*b*) *S*s develop a strategy which shortcuts the process of stimulus generation. Possibly, *S*s categorize the single-target stimulus with a code. Should an identity mismatch occur between the single-target member and the probe, the code-matching operation, which may be initiated parallel to the identity-matching operation, is employed. This coding strategy could also be used for dual targets. However, spontaneously introspective comments from *S*s uniformly indicated that they were impelled by the physical presence of both members of the positive set, dual targets, to perform an exhaustive identity match for the right and left members of the set prior to making a difference judgment. Thus dual- and single-target matching efficiency may both suffer at the shortest ISI, but for different reasons. For dual targets, *S*s may be less able to shift attention between visual images of both stimuli before the probe arrives. For single targets, the coding operation may be less well implemented by the time the probe arrives. An alternative hypothesis is that *S*s ignored the dual-target information, selected one member of the target set, and performed a set of mental operations identical to those for single targets. This hypothesis is not confirmed by these data; matching times did not differ significantly for the Dual-IL and

Dual-IR conditions, and Dual-IL and Dual-IR times were significantly longer than those for the Single-IS condition.

While introspective comments by *S*s regarding an exhaustive, serial, double identity-matching operation for dual targets must be interpreted with caution, they are not inconsistent with Posner and Boies' (1971) argument that the target stimulus sensitizes *S*s to identical stimuli and serves as a model for search. Further, introspective comments are consistent with the finding that identity matching takes priority over category matching, regardless of matching instructions. They are also consistent with the finding that Dual-IL and Dual-IR conditions required significantly greater amounts of time than matching for the single-target IS category. The amount of time required to complete IS matching plus NIS and DS matching for single targets was about the same as that required to complete IL plus IR matching for dual targets. Although this finding does not necessarily indicate that the former operations were done in parallel while the latter were executed in a strictly serial fashion, it does suggest that considerably more overlap in time is possible for dissimilar operations, identity (NIS), and category matching (DS) than for highly similar operations such as "double identity matching" (Dual IR and IL). Further studies are necessary to determine the specific nature of preprobe activity for material varying in spatial orientation and to test the proposition that highly similar operations have to be carried out in series, or in parallel with mutual interference (Posner & Boies, 1971), while dissimilar operations may be conducted in parallel without interference.

REFERENCES

BOIES, S. J. Rehearsal of visual codes of single letters. Unpublished master's thesis, University of Oregon, 1969.

BOIES, S. J. Memory codes in speeded classification tasks. Unpublished doctoral dissertation, University of Oregon, 1970.

COLE, J. W. L., & GRIZZLE, J. E. Applications of multivariate analysis of variance to repeated measurements experiments. *Biometrics*, 1966, **22**, 810–828.

DAINOFF, M. J. Time course of visual and auditory encoding. *Journal of Experimental Psychology*, 1970, **86**, 214–224.

MORTON, J. Interaction of information in word recognition. *Psychological Review*, 1969, **76**, 165–178.

POSNER, M. I. Abstraction and the process of recognition. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation*, New York: Academic Press, 1969.

POSNER, M. I., & BOIES, S. J. Components of attention. *Psychological Review*, 1971, **78**, 319–408.

POSNER, M. I., BOIES, S. J., EICHELMAN, W. H., & TAYLOR, R. L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology*, 1969, **79** (1, Pt. 2).

POSNER, M. I., & Keele, S. W. Decay of visual information from a single letter. *Science*, 1967, **158**, 137–139.

POSNER, M. I., & Keele, S. W. Time and space as measures of mental operations. Paper presented at the meeting of the American Psychological Association, Miami Beach, September 1970.

POSNER, M. I., & TAYLOR, R. L. Subtractive method applied to separation of visual and name components of multi-letter arrays. *Acta Psychologica*, 1969, **30**, 104–114.

SHEPARD, R. N., & METZLER, J. Mental rotation of three-dimensional objects. *Science*, 1971, **171**, 701–703.

STARMER, C. F. A multivariate analysis program for biomedical research. Paper presented at the sixth annual Southeastern meeting of the Association for Computing Machinery and National Meeting of Biomedical Computing, Durham, N. C., June 1967.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received December 17, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 184–188

GENERAL AND SPECIFIC FACTORS IN THE INTERSENSORY TRANSFER OF FORM[1]

JEFFREY L. CLARK,[2] JOEL S. WARM, AND DONALD A. SCHUMSKY

University of Cincinnati

This study assessed the relative contributions of specific and nonspecific components to intersensory transfer between vision and touch. A paired-associate paradigm was used in which visual metric figures and their tactual analogs served as stimuli, and familiar adjectives were the responses. Positive intersensory transfer, characterized by symmetry across modalities was obtained. The contribution of nonspecific learning to this effect was negligible. Intersensory transfer was found to be less efficient than the empirically determined maximum level of intrasensory transfer possible in this task.

Intersensory transfer refers to the effect of initial learning in one sensory modality on later learning in another modality (Gibson, 1969). A paired-associate paradigm has frequently been used to explore intersensory transfer effects. In an early study, Gaydos (1956) had *S*s learn associations between nonsense patterns, presented visually or tactually, and verbal labels. Afterward, *S*s shifted modalities and relearned the stimulus–response pairs. Positive transfer was obtained—significantly fewer trials were required for relearning in the second modality than in the training mode. Subsequent studies by Eastman (1967) and by Walk (1965) have also demonstrated visual-tactual transfer using the paired-associate technique.

Although intersensory transfer clearly occurs, the mechanisms for such transfer remain to be identified. Certainly, perceptual models, based upon stimulus generalization or overlap in the coding of stimulus elements, and mediational models, based upon some form of learned association between the two modes of experience, might afford possible explanations (Ellis, 1969; Gibson, 1969; Holmgren, Arnoult, & Manning, 1966). However, an appeal to nonspecific factors such as warm-up or learning-to-learn (Ellis, 1969) is also plausible. In this regard, Holmgren et al., using acoustic and visual stimuli, reported that performance on a transfer task in one modality was facilitated by initial training in another modality although the stimulus–response units in both phases of the experiment were dissimilar. These investigators concluded that nonspecific factors may indeed have been involved in the audiovisual transfer that they observed.

The studies of visual-tactual transfer by Eastman (1967), Gaydos (1956), and Walk (1965) were not designed to permit a similar analysis of the components of transfer. Consequently, these studies invite replication under conditions that would permit assessment of the role of general and specific factors in transfer of training between the visual and tactual modes. Such assessment represented one goal of the present experiment.

In addition to transfer of training, intermodal relations have also been studied by matching procedures which require *S*s to select comparison patterns in one modality that are identical to standard patterns presented in a different mode (Gibson, 1969). Performance on this type of cross-modal task exceeds chance expectations, but it is generally poorer than performance on a similar intramodal matching task, particularly in regard to visual-visual and visual-tactual comparisons (Foulke & Warm,

---

[1] This report is based, in part, upon a thesis submitted by the first author to the Graduate School of the University of Cincinnati in partial fulfillment of the requirements for the MA degree. The research was supported by the Institute of Space Sciences of the University of Cincinnati under NASA Grant No. NGL 36-004-014. The authors are indebted to W. N. Dember, H. C. Ellis, A. F. Grasha, and R. P. Honeck for their helpful comments.

[2] Requests for reprints should be sent to Jeffrey L. Clark, Department of Psychology, Pfeiffer College, Misenheimer, North Carolina 28109.

1968; Gibson, 1969; Lobb, 1965). This suggests that in cross-modal matching, the interchange of information between the senses is not complete. To date, the identification of a similar performance decrement with regard to transfer of training has been neglected. Accordingly, a second aspect of the present study was to compare the level of intermodal transfer to an empirically determined maximum level of intramodal transfer. In this way, any decrement in transfer associated with crossing sensory channels could be identified.

METHOD

Subjects.—Sixty *S*s, 25 men and 35 women, served in the experiment. They ranged in age from 18 to 48 with a mean age of 20 yr. All *S*s were from psychology classes at the University of Cincinnati and were naive as to the problem under investigation.

Stimulus items.—The stimuli were sixteen visual metric figures and their tactual analogs. Metric figures resemble histograms and are generated by selecting column heights, according to a predetermined sampling rule, in a symmetrical Row × Column matrix. All figures were constructed on the basis of an underlying 4 × 4 matrix and were generated by random sampling of column heights with replacement. Figures generated by this sampling rule are known as random metric figures (Fitts, Weinstein, Anderson, & Leonard, 1956).

The visual figures were constructed by using small solid squares produced by a special typewriter key. This key produced .25 × .25 cm. squares; each square represented one cell of the metric figure. The figures were centered on 10.24 × 15.36 cm. white index cards. They were coated with black ink, which served to remove graininess, to provide clear contours, and enhance contrast with the white background.

The tactual or haptic analogs of the visual figures were punctiform patterns, embossed on aluminum by means of a specially constructed braille slate and mounted on Masonite blocks .64 cm. thick, and 4.48 cm. on a side. Vertical and horizontal spacing between the center of adjacent dots was .24 cm. This value is standard for the braille code. The figures so constructed were similar to those used in a previous series of studies (Foulke & Warm, 1967; Warm, Clark, & Foulke, 1970; Warm & Foulke, 1968).

Responses.—The response terms were 16 five-letter adjectives with a frequency of occurrence in ordinary use of 100 per million words as indexed by Thorndike and Lorge. Each adjective was paired with a stimulus form at random. The resulting 16 stimulus–response pairs were then divided into two groups of eight paired-associates designated Stimulus–Response Set 1 (SR-1) and Stimulus–Response Set 2 (SR-2).

Apparatus.—A two-channel tachistoscope was used for visual presentation. The stimulus figures subtended a visual angle of .74°. The luminance of the adapting and exposure fields of the tachistoscope was 1.4 ftl., and *S*s remained light adapted throughout the experimental session.

Stimuli were presented for tactual inspection in a partially enclosed box placed on a table between *S* and *E*. This arrangement prevented visual inspection of tactual stimuli by *S*. The Masonite block containing the stimulus figure was fitted into an appropriate jig in the floor of the apparatus. The *S* gained access to the tactual stimulus by placing his preferred hand through the curtained front of the box, and he inspected the stimulus actively with his index finger. (A more complete description of the tactual apparatus can be found in Foulke and Warm, 1967.) In both the visual and tactual conditions, all figures were presented with base horizontal and distinctive contour pointing upward.

Procedure.—Each *S* was run individually in a session that lasted approximately 35 min. Experimental sessions were divided into training and transfer phases; *S*s were required to learn by anticipation a list of eight paired-associates to a criterion of *one* errorless trial in both phases. A trial was defined as one presentation of all eight items in the list. The sequence of item presentations was varied at random during each trial for all *S*s.

In order to minimize response learning, the eight adjectives appropriate to a given phase of the experimental session were displayed in a prominent position on the relevant apparatus. As pointed out by Gaydos (1956), the equivalence of sensory contact is a critical issue in comparing visual and tactual pattern recognition; encoding time for vision is generally much shorter than that necessary for touch. In view of this problem, no time limits were set for either visual or tactual inspection of the stimuli. This procedure is consistent with that followed by Gaydos. After inspecting a form, *S* responded with one of the adjectives and then was informed immediately by *E* of the correct response to the item. A 10-sec. interitem interval was maintained throughout the experiment. The *S*s were given a 1-min. rest period between the training and transfer phases of the session.

Experimental design.—Ten *S*s were assigned at random to each of six experimental groups. The groups were defined by the sensory modality and stimulus–response sets used in each phase of the experiment. These groups are outlined in Table 1.

Groups T(SR-1):V(SR-1) and V(SR-1):T(SR-1), in which sensory modalities differed during the training and testing phases but the stimulus–response set remained the same, were designed to permit assessment of the presence of specific intermodal transfer. Groups T(SR-2):V(SR-1) and V(SR-2):T(SR-1), in which both modalities and stimulus–response sets differed in both phases of the experiment, were designed to permit assessment of nonspecific learning effects in intermodal transfer. The remaining groups, T(SR-1):T(SR-1) and V(SR-1):V(SR-1), in which both stimulus–re-

TABLE 1

SENSORY MODALITIES AND STIMULUS–RESPONSE SETS EMPLOYED DURING TRAINING AND TRANSFER PHASES FOR ALL GROUPS

Group	Training	Testing	S–R sets during both phases
T(SR-1): V(SR-1)	Tactual	Visual	Same
V(SR-1): T(SR-1)	Visual	Tactual	Same
T(SR-2): V(SR-1)	Tactual	Visual	Different
V(SR-2): T(SR-1)	Visual	Tactual	Different
T(SR-1): T(SR-1)	Tactual	Tactual	Same
V(SR-1): V(SR-1)	Visual	Visual	Same

sponse sets and sensory modalities remained invariant in both experimental phases, were control groups designed to (*a*) provide a base for evaluating specific and nonspecific transfer effects, and (*b*) permit a measure of the maximum intrasensory transfer possible under the conditions of this study. These two groups also provided a base for identifying any decrement in transfer associated with changing sensory channels.

RESULTS

Original learning.—Before an adequate assessment of intermodal transfer could be made, it was necessary to establish the equivalence of the two stimulus–response sets used in initial learning. Consequently, a 2 × 2 factorial combination of stimulus–response sets and sensory modalities was derived from the training data of Groups **V(SR-1)**:V(SR-1), **T(SR-1)**:T(SR-1), **V(SR-2)**:T(SR-1), and **T(SR-2)**:V(SR-1). The mean number of trials to criterion with SR-1 were 4.8 and 8.0 for vision and touch, respectively. With SR-2, the corresponding values were 4.8 for vision and 8.1 for touch. An analysis of variance of the data indicated that no significant differences in learning were associated with the stimulus–response sets ($F < 1$). Moreover, while initial learning was superior for vision as compared to touch, $F\ (1,\ 36) = 57.48$, $p < .01$, the absence of a significant Modality × Sets interaction ($F < 1$) indicated that initial learning with each stimulus–response set was similar within each of the sensory modalities employed.[3]

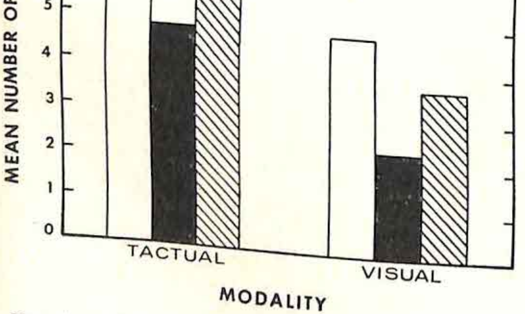


FIG. 1. Mean number of trials to criterion during initial learning and intermodal transfer for both sense modalities.

Intermodal transfer.—The relative contributions of specific and nonspecific factors to intermodal transfer in both modalities were assessed by comparing the training data of Groups **V(SR-1)**:V(SR-1) and **T(SR-1)**:T(SR-1) to the specific intermodal transfer data of Groups T(SR-1):**V(SR-1)** and V(SR-1):**T(SR-1)** as well as to the nonspecific transfer data of Groups T(SR-2):**V(SR-1)** and V(SR-2):**T(SR-1)**. The mean number of trials to criterion for these training and transfer conditions are presented in Fig. 1.

The figure reveals that overall performance was superior with visual as compared to haptic stimulation. For both modalities, the specific transfer groups reached criterion in roughly 50% fewer trials than the original training groups, while the performance of the nonspecific transfer groups was generally similar to that of *S*s experiencing the original training conditions. The data were subjected to a 2 × 3 analysis of variance in which sensory modality was one dimension and type of experience (original learning, specific transfer, and nonspecific transfer) was the other. Significant differences were noted between sense modes, $F(1, 54) = 67.14$, $p < .01$, and between types of experience, $F(2, 54) = 18.76$, $p < .01$. The Modality × Experience interaction lacked significance ($F < 1$) indicating that the differences between training and transfer conditions were symmetrical across the two sensory modalities.

---

[3] The training data of Groups **V(SR-1)**:T(SR-1) and **T(SR-1)**:V(SR-1) were omitted from this analysis in order to retain an equal number of *S*s in each cell of the factorial design. However, the mean number of trials to criterion for initial learning in these groups were nearly identical to those of the corresponding SR-1 groups used in the analysis. These values were 4.7 for vision and 8.2 for touch.

To examine further the effects of type of experience on transfer, a Duncan multiple-range test was performed using a protection level of .05 (Edwards, 1968). This test revealed that the specific intermodal transfer condition required significantly fewer trials to criterion than either the original learning condition or the nonspecific intermodal transfer condition while the latter two did not differ significantly from each other. Taken together, the results of the analysis of variance and the associated Duncan test indicate that positive intermodal transfer was obtained in this experiment, that such transfer was symmetrical across modalities, and that the contribution of nonspecific factors to this effect was negligible.

Intermodal versus intramodal transfer.— Two groups, namely, V(SR-1):V(SR-1) and T(SR-1):T(SR-1), learned the paired-associate task to criterion during training and then, continued with the identical stimulus–response units during the transfer phase of the experiment. Every *S* in these conditions correctly named all of the stimulus figures on the first transfer trial. This perfect performance of one trial was used as the base line for evaluating any performance decrement associated with transition between the sensory modalities.

Accordingly, the efficiency of cross-modal relative to intramodal transfer was assessed by comparing the transfer stage performance of Groups V(SR-1):**V(SR-1)** and T(SR-1):**T(SR-1)** to the transfer data of the specific cross-modal Groups T**(SR-1)**: **V(SR-1)** and V(SR-1):**T(SR-1)**, respectively. The mean number of trials to criterion for both intramodal conditons was, of course, 1. By contrast, the mean number of trials to criterion was 2.3 in cross-modal transfer from touch to vision and 4.8 in transfer from vision to touch. Direct statistical tests of the differences between the cross-modal and intramodal transfer conditions were precluded by the lack of variance in the intramodal data. However, in order to insure that these differences were statistically reliable, the 5% confidence intervals were computed for the means of the cross-modal conditions. For both modalities, these intervals excluded the mean of the intramodal groups. Thus, the data reveal a statistically significant performance decrement associated with the transition between the sensory modalities.

DISCUSSION

This study was designed to assess the role of general and specific factors in intermodal transfer between vision and touch. Positive transfer effects that were symmetrical across modalities were noted and nonspecific components were not found to contribute substantially to this result.

An important methodological feature in the present investigation was the use of a self-paced paired-associate paradigm. This procedure was employed in order to allow equivalent stimulus encoding in a situation where, between sensory modalities, effective encoding times must and do vary. It might be argued, however, that the self-determined time procedure introduced a source of confounding into the trials to criterion evaluation of transfer used here. The apparently large contribution of specific relative to nonspecific transfer components may be due to the fact that in the specific transfer conditions, *S*s simply took longer inspection times before responding and thus, required fewer trials to learn the transfer task. While this is a possibility, there is strong reason to believe that it is remote and that if inspection time differences did occur, they would have been in the direction of longer times for the nonspecific transfer *S*s. Zacks (1969) has reported that when *S*s are given free inspection time in a paired-associate paradigm, there is a significant positive correlation between study time and item difficulty. In addition, inspection times after errors tend to be twice as long as after correct responses. Clearly, long inspection times tend to be associated with increased not decreased difficulty in self-paced paired-associate learning. All in all, it appears that the positive intersensory transfer noted in this study can be most parsimoniously interpreted in terms of specific stimulus–response relations and not a methodological artifact.

As indicated earlier, current attempts to explain intersensory transfer have appealed to mediational and perceptual models. Although the present investigation was not designed to permit an empirical test of these approaches, the data have meaning for both of them. The pattern of performance decrement in crossing sensory channels that has been noted in this study must be accounted for in any attempt to provide a thorough understanding of the phenomenon of intersensory transfer. Among other things, such a decrement could result from (*a*) insufficient strength of implicit associations or response produced cues (cf. Ellis, 1969), (*b*) incomplete encoding of invariant features of stimulation common to the modalities involved, or (*c*) an inability to utilize the extracted information completely in joining the paired associates in the cross-modal task. The degree to which factors such as these underlie the performance decrement in the interchange between modalities should be assessed in future experiments.

In addition to the examination of intermodal transfer effects, comparisons between vision and touch in regard to rates of initial learning are also of interest. More specifically, the levels of initial learning for visual and haptic stimulation were essentially similar in the early study reported by Gaydos (1956), whereas initial learning was superior with visual as compared to tactual stimuli in the present investigation. To a naive *O*, the pattern detail of the punctiform figures used here was probably less discriminable than that of the tactual figures used by Gaydos; the figures described in his report were considerably larger and had more gross contour variation. The increased difficulty of the haptic discriminations required in this study probably enhanced visual-tactual differences in initial learning. This interpretation is consistent with Walk's (1965) finding that differences in the learning of forms with vision and touch are highly dependent upon the characteristics of the forms employed.

REFERENCES

Eastman, R. The relative crossmodal transfer of a form discrimination. *Psychonomic Science*, 1967, **9**, 197–198.

Edwards, A. L. *Experimental design in psychological research.* (3rd ed.) New York: Holt, Rinehart & Winston, 1968.

Ellis, H. C. Transfer: Nature, measurement and fundamental processes. In M. H. Marx (Ed.), *Learning: Processes.* New York: Macmillan, 1969.

Fitts, P. M., Weinstein, M., Anderson, N., & Leonard, J. A. Stimulus correlates of visual pattern recognition: A probability approach. *Journal of Experimental Psychology*, 1956, **51**, 1–11.

Foulke, E., & Warm, J. S. Effects of complexity and redundancy on the tactual recognition of metric figures. *Perceptual and Motor Skills*, 1967, **25**, 177–187.

Foulke, E., & Warm, J. S. Intra–intermodal identification of visual and tactual patterns. Paper presented at the meeting of the Psychonomic Society, St. Louis, November 1968.

Gaydos, H. F. Intersensory transfer in the discrimination of form. *American Journal of Psychology*, 1956, **19**, 175–187.

Gibson, E. J. *Principles of perceptual learning and development.* New York: Appleton-Century-Crofts, 1969.

Holmgren, G. L., Arnoult, M. D., & Manning, W. H. Intermodal transfer in a paired-associates learning task. *Journal of Experimental Psychology*, 1966, **71**, 254–259.

Lobb, H. Vision versus touch in form discrimination. *Canadian Journal of Psychology*, 1965, **19**, 175–187.

Walk, R. D. Tactual and visual learning of forms differing in degrees of symmetry. *Psychonomic Science*, 1965, **2**, 93–94.

Warm, J. S., Clark, J. L., & Foulke, E. Effects of differential orientation on tactual pattern recognition. *Perceptual and Motor Skills*, 1970, **30**, 987–994.

Warm, J. S., & Foulke, E. Effects of orientation and redundancy on tactual perception of form. *Perceptual and Motor Skills*, 1968, **27**, 83–89.

Zacks, R. T. Invariance of total learning time under different conditions of practice. *Journal of Experimental Psychology*, 1969, **82**, 441–447.

(Received December 20, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 189–194

STIMULUS SELECTION AND MEANINGFULNESS AT DIFFERENT STAGES OF PAIRED-ASSOCIATE LEARNING[1]

FRANKLIN M. BERRY,[2] CHARLES E. JOUBERT,[3] AND ALFRED A. BAUMEISTER

University of Alabama

Redundant-stimulus paired associates (PAs) consisting of word-trigram stimuli and digit responses (e.g., CAT + BWD:0) were practiced to different degrees of PA learning (either 1, 2, 3, 5, 1 Perfect, or 5 Perfect trials) by 12 subgroups of college students. Six of the subgroups were tested for intercue associations and 6 were tested for cue selection. The *S*s were observed to rely almost exclusively upon the word cues at all stages of learning. Prior to overlearning (5 Perfect trials), there was no evidence for the existence of either trigram:response or trigram:word associations. Following overlearning, however, a parallel increase was observed in the availability of both trigram:response and trigram:word associations. It was concluded that the hypothesis of indirect association of lower meaningfulness (M) cues and responses (i.e., trigram:word:response associations) was sufficient to "explian" the parallel rise in availability of apparent trigram: response and trigram:word associations.

The relationship between stimulus meaningfulness (M) and stimulus selection in paired-associate (PA) learning by college students is a relatively well-established phenomenon. College *S*s almost always select the more meaningful of two cues for use as their *functional stimulus* during PA learning. That this relationship is a general one is attested to by the observation that two different stimulus redundancy paradigms have yielded similar findings. When low-M trigrams were combined with colors to form compound cues, *S*s were observed to rely almost exclusively upon the colors for recall of the responses (e.g., Houston, 1967; Underwood, Ham, & Ekstrand, 1962). Similarly, when low-M trigrams were combined with words to form compound stimuli, *S*s were observed to rely primarily upon the word cues (e.g., James & Greeno, 1967; Spear, Ekstrand, & Underwood, 1964).

Nevertheless, results from both the color-trigram and word-trigram redundancy paradigms have shown that the lower M cues (trigrams) usually acquire at least some cue value during PA learning, although low-M trigrams are not selected with nearly the same frequency as words. However, the repeated finding that *S*s do attend to the trigrams raises the question of just how these cues enter into associative learning.

Two alternative notions were offered by Underwood et al. (1962) to account for the role played by low-M trigrams during PA acquisition: (*a*) The trigrams are directly associated with their respective responses such that trigram:response associations are formed, and (*b*) the trigrams are directly associated with the more meaningful elements of the compounds and these, in turn, are directly associated with the responses. Inasmuch as experimental evidence has accumulated for the existence of cue:cue associations in both the color-trigram (Birnbaum, 1966) and the word-trigram (Spear et al., 1964) paradigms, the choice between the direct association and the indirect association hypotheses is no simple matter.

A methodological difficulty inherent in testing the indirect-association hypothesis must be recognized. The most straightforward approach to inferring the existence of cue:cue associations, viz., *S*'s ability to match up cues presented together as a compound during PA learning, is not sufficiently analytical to permit unequivocal inferences about the availability of intercue associations. Spear et al. (1964)

---

[1] This research was supported, in part, by United States Public Health Service Research Grant HD 02588. The third author was the principal investigator.

[2] Requests for reprints should be sent to Franklin M. Berry, Department of Psychology, Columbus College, Columbus, Georgia 31907.

[3] Now at Southeastern Louisiana College, Hammond, Louisiana 70401.

were first to point out that successful cue:cue matching could be response mediated. That is, if *S*s form separate associations between each of the stimulus components and the correct response, then it is possible to match appropriate components by linking together a "forward" and a "backward" association involving a common response (S_1:R:S_2 or S_2:R:S_1). Steiner and Sobel (1968) devised direct tests of such chaining in a color-word, redundant-stimulus PA investigation. While they found no evidence for the S_1:R:S_2 chaining mechanism, they did find clear support for the availability of cue:cue (color:word) associations.

The goal of the present study was to provide further information concerning the acquisition of both interstimulus (S–S) associations and stimulus–response (S–R) associations during redundant-cue PA learning.[4] The word-trigram redundancy paradigm was employed, and both S–S and S–R associations were measured at different degrees of PA learning. The S–S associations were measured by matching tasks and, for strict comparability of S–S and S–R conditions, S–R associations were likewise measured by matching tests. It also should be noted that previous investigators have recognized the importance of examining the variable of degree of learning in the color-trigram (Houston, 1967) and word-trigram (James & Greeno, 1967) redundancy paradigms. However, these studies have focused only on cue:response associations, and they have not produced entirely consistent findings (see Berry, Joubert, & Baumeister, 1971).

METHOD

General procedure.—Both S–S and S–R associations were measured at six stages of PA learning: 1, 2, 3, 5, 1 Perfect or 5 Consecutive Perfect trials. The S–S associations were measured in two "directions" (word:trigram and trigram:word) at each stage of learning. Since a given *S* was tested in only one direction, 12 subgroups were required to measure S–S associations. The effects of differential availability of words and trigrams were controlled by providing *S*s cards bearing all word cues or all trigram cues, respectively. For example, *S*s were handed the card containing all the trigrams when tested for word:trigram associations. The measurement of cue:response associations was accomplished with a within-*S*s design: a given *S* practiced the pairs until he reached a prescribed criterion, and then he was tested for response recognition to: (*a*) words and trigrams presented together as in PA learning (control), (*b*) words only, and (*c*) trigrams only. The *S*s tested for cue:response associations were provided a card containing all the response digits during the cue-selection tests. The order of presenting the cue-selection tests was controlled by use of a balanced Latin square at each stage of learning.

Subjects.—The *S*s were 114 students enrolled at either the University of Alabama or Southeastern Louisana College. They were randomly assigned to treatments by means of a random number table (all protocols were randomized before the experiment was started).

Apparatus, lists, and materials.—The apparatus consisted of a Kodak Carousel slide projector, a projection screen, and associated timers and relay circuitry for automatically timing slide presentation intervals. The basic list of word-trigram PAs was formed by randomly combining high-M words (Noble, 1961), low-M trigrams (Krueger, 1934; Witmer, 1935), and single digits. The trigrams, as indexed by the number of repeated letters, were of relatively high formal similarity. Similar trigrams were used to minimize the effectivenss of *single-letter* cue-selection strategies. The list consisted of: CAT + BWD:0, CAR + BXD:1, BAG + DWK:2, NEW + XYH:3, MAN + GVS:4, BOY + DXH:5, DIE + NXQ:6, BED + GWS:7, DAY + NHQ:8, and NOW + XFH:9. The three transfer test lists which were derived from the OL list were: (*a*) control list—both the words and trigrams were presented; (*b*) word list; and (*c*) trigram list. The response terms were not present on any of the transfer slides.

Procedure.—Each *S* was read a standard set of instructions for study-test PA learning. (These instructions were designed to give equal emphasis to both the trigrams and the words.) Study and test periods were alternated and, for each, a 2-sec. rate was employed. On test trials, *S* was requested to call out the digit previously paired with each redundant stimulus. Four random orders of study slides and four different random orders of test slides were used. An equal number of *S*s was started on each of the four study slide orders. For study and test trials, the compound stimuli were presented always on the left side of the slide, one above the other (the words appeared above the trigrams for half of the stimuli and the trigrams appeared above the words for the other half). On successive trials the position of each stimulus component was alternated, e.g., above, below, above, below. For any given trial, however, a particular stimulus component was presented in the same position for "study" as for "test." The *S*s practiced the PAs to their assigned performance criterion, and then they entered the transfer phase.

---

[4] The present usage of S–S as opposed to S–R associations is congruent with the operations used to measure intercue and cue-response associations, respectively; it in no way reflects a presumption as to the ultimate nature of what is learned by *S*.

The transfer phase was begun immediately following PA learning; each *S* was read the appropriate transfer instructions. The *S*s tested for cue:response associations were shown the words alone, the trigrams alone, and both the words and trigrams together in a counterbalanced fashion. Three different orders of the lists were used to measure these associations. An equal number of *S*s ($N = 3$) was tested with each order. None of these orders duplicated those used elsewhere in the experiment. The *S*s who were tested for the development of cue:cue associations were shown either the word stimulus alone ($N = 5$) or the trigram stimulus alone ($N = 5$) and were required to match it with the other stimulus component. The transfer tasks were *S*-paced and no feedback was given. The *S*s were encouraged but not required to guess.

RESULTS

Original learning.—Several analyses were conducted to test the assumption that the different subgroups in the design were about equal in PA learning ability. As all of these analyses led to the same conclusion, only two are reported. The first analysis involved only the subgroups to be tested for S–S associations (either trigram to word or word to trigram), and it was based on the number of correct responses that each *S* achieved on the first trial of PA learning. There were a total of 12 subgroups in all, 2 subgroups (each $N = 5$) at each of 6 possible levels of PA learning. The means for these subgroups ranged from 1.2 to 4.0. A 2 × 6 (Type of Subgroup × Degree of Learning) analysis of variance indicated that there were no significant sources of variance (all $F\text{s} < 1.0$). Since the pairs of subgroups from each of the 6 levels of PA learning were not different, their data were combined to yield 6 subgroups (each $N = 10$) and these were compared with the 6 subgroups to be tested for S–R associations (each $N = 9$). The means of correct responses for the 6 S–R subgroups ranged from 1.3 to 2.9. The results of a second 2 × 6 (Type of Subgroup × Degree of Learning) analysis of variance supported the notion that the S–S and S–R subgroups which were to be contrasted in transfer performed comparably in PA learning (all $F\text{s} \leq 1.1$).

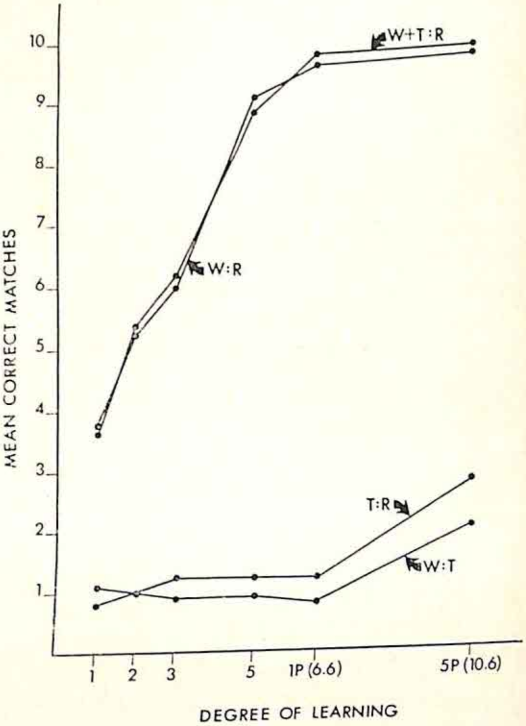


FIG. 1. Mean correct matches for the S–S (W:T) and S–R subgroups (W+T = compound cue; W = word cue; T = trigram cue) when scoring was by a single criterion.

Transfer.—The first analysis involved the 12 S–S subgroups, 2 from each of the 6 levels of PA learning. A 2 × 6 (Direction of Recall × Degree of Learning) analysis of variance was performed on the number of correct S–S matches achieved by each *S*. The means for this analysis ranged from .4 to 2.6. As the analysis revealed that direction of recall was not a significant variable, $F\ (1,\ 48) = 1.4,\ p > .05$, a decision was made to pool the matching data for each of the pairs of subgroups trained to the same level of PA learning. The combined means of correct word-trigram (W:T) matches are shown in Fig. 1. (It should be noted that the designation W:T is an arbitrary convenience as T:W and W:T matching data are both represented in these means.)

Figure 1 also presents the mean numbers of correct matches for the S–R subgroups presented compound cues (W+T:R), words alone (W:R), and trigrams alone (T:R). The curves for the S–R subgroups tested under single-cue conditions (W:R or T:R) reflect *single-criterion* scoring. That is, a correct response to an individually presented cue (word or trigram) was the sole criterion for inferring that that cue was *S*'s functional stimulus. Inspection of Fig. 1 suggests that the words were most often selected as the functional cues for PA learning. Indeed, the W+T:R and W:R curves are indistinguishable across trials. Figure 1 also shows another typical result for redundant-cue PA learning, namely, that the trigrams account for a small number of correct responses. Finally, it can also be seen that the curve for the performance of the S–S subgroups on the word-trigram matching tests is strikingly similar in form to the curve for the S–R subgroups on the trigram-response test.

Two separate analyses of variance were used to test the trends suggested in Fig. 1. In the first, involving only the S–R subgroups, correct responses were analyzed in a 3 × 6 (Type Cue × Degree of Learning) mixed analysis of variance. Both main effects as well as their interaction proved significant: (*a*) type cue, F (2, 96) = 218.0; (*b*) degree of learning, F (5, 48) = 16.6; and (*c*) interaction, F (10, 96) = 5.3, all $p\text{s} < .01$. Simple-effects analysis showed that no significant differences existed between the W+T:R and W:R conditions at any stage of learning, and, also, that both were superior to the T:R condition at all learning stages. Similarly, degree of learning was significant with respect to each cue condition, although the effect was far less pronounced for the T:R condition, F (5, 48) = 2.0, $p < .10$.

A second analysis was performed to compare the performances of the W:T and T:R groups at each stage of learning. Only the degree of learning main effect was significant: F (5, 102) = 3.3, $p < .01$. Subsequent t tests showed that performance following 5 Perfect trials was significantly greater than that following all other stages of practice (all $p < .01$); no other comparisons yielded significant differences.

DISCUSSION

These results confirm the now common observation that college students tend to select the more meaningful of two redundant-cue elements for completion of the associative stage of PA learning. As in earlier studies using either the color-trigram or word-trigram paradigms, the higher M cues (words) were found to be completely effective stimuli for associative learning.

But what of performance based on the trigrams? As indicated in Fig. 1, there is no evidence for the availability of T:R associations prior to overlearning; at all other stages of practice, T:R matching can be accounted for on the basis of chance. A similar conclusion appears to hold for the W:T data (see Fig. 1). In effect, this means that the indirect hypothesis probably received only one reliable "test," the 5 Perfect trial comparison. In this test, the predicted equality of apparent T:R associations and W:T associations was upheld. Another finding, which is also congruent with the indirect hypothesis, emerges from a consideration of the transfer test performance of the S–R subgroup receiving extended PA practice. For this group, the indirect hypothesis must predict that any *S* who was able to complete a correct T:R match for a PA pair must also be able to complete a correct W:R match for the same PA pair. This was in fact what occurred (See Fig. 1, recalling that the different S–R associations were measured within the same *S*).

The present findings tend to support the earlier work of James and Greeno (1967). Both studies show that trigram selection is limited to pairs previously learned as W:R pairs and also that such selection tends to occur only after all pairs in the list have been learned. In addition, the present findings suggest that the apparent increase in the availability of T:R associations following overlearning occurs as a result of intercue mediation, in support of the indirect association hypothesis. Although the present overlearning findings of a parallel rise in T:R and W:T associations can be "explained" parsimoniously by the indirect association hypothesis, the possibility remains that *S*s of the present experiment formed independent trigram:response and word:trigram associations.

In an effort to demonstrate some generalities concerning cue selection, we shall compare our estimates of T:R availability following precriterion, criterion, and postcriterion levels of learning with earlier ones from previous word-trigram studies.[5] These systematic comparisons provide further information about the role played by trigram cues in word-trigram PA learning. We selected the dependent measure of percent correct T:R association for making comparisons between the different studies. In all cases, we calculated these measures from information presented in each experiment.

Precriterion estimates.—First, consider estimates of the likelihood that T:R associations are formed during precriterion trials. The only relevant estimates are those of James and Greeno (1967; Exp. I; Exp. II) and ours. In both cases, there were only about 10% correct T:R matches. It would appear that *S*s rarely attend to trigrams during precriterion trials.

Criterion estimates.—Now we consider estimates that T:R connections are available following list mastery. Our estimate of T:R availability, 11% (following 1 Perfect trial; see Fig. 1), is lower than those from most relevant studies (Cohen & Musgrave, 1964; Harrington, 1969; James & Greeno, 1967, Exp. I, II, III; Lovelace, 1968; and Spear et al., 1964, Exp. 1). The estimates of T:R availability obtained in these studies extend from about 5% to about 50% of the pairs. However, most of the estimates fall within a range of about 18%–28% of the pairs. Why are all these values so discrepant? An examination of the word-trigram literature suggests that the answer to this question probably involves the different methods which have been used to present the word-trigram cues during PA learning: (*a*) The highest T:R estimate (Cohen & Musgrave, 1964; their LH-L condition) occurred when all trigrams in the list were consistently presented to the left of their word cues. (*b*) The lowest estimate (James & Greeno, 1967, Exp. I) occurred when all trigrams in the list were consistently presented to the right of their word cues. (*c*) The middle-range estimates (about 25% of the pairs) occurred in experiments in which the left–right orientation of the trigrams and words was balanced within the PA list.

This does not appear to be the whole story, however. Our criterion estimate of T:R availability is also lower than the middle-range estimate of Spear et al., 1964, Exp. 1; but, neither the trigram nor the word was presented in a consistent left–right orientation in these studies. Rather, the positional orientation of each stimulus element was varied on each successive trial (e.g., above, below, above, below, and so on). A further comparison of the two studies reveals that they differed most conspicuously in the characteristics of the trigram cues. In the present work, trigrams with many repeated letters were used, but in the earlier experiment, trigrams with few repeated letters were employed. It is highly probable that our criterion estimate of T:R availability is lower than that of Spear et al. because the opportunities for single-letter cue selection were minimized in the present investigation.

Postcriterion estimates.—Now consider estimates that T:R connections are available following overlearning. Only one previous word-trigram study has been concerned with the effects of overlearning on T:R availability, that of James and Greeno (1967, Exp. I, II, III). Their estimates of T:R availability following 10 practice trials beyond list mastery (1 Perfect trial) are only roughly comparable to our own estimate of 25%. All four experiments are, however, in close agreement concerning the observation that T:R availability increases as a function of overlearning.

In summary, a comparison of the present results with the existent word-trigram literature indicates that: (*a*) for precriterion levels of learning, *S*s do not appear to process lower M trigrams for associative learning; and (*b*) for criterion and postcriterion levels of learning, the probability that *S*s process trigrams seems to depend on both the left–right placement of the word and trigram cues and on the possibilities for selecting single letters from the trigrams for use as effective stimuli. Similarly, all existing studies indicate that overlearning leads to increased processing of the trigram cues.

REFERENCES

BERRY, F. M., JOUBERT, C. E., & BAUMEISTER, A. A. Single-letter cue selection and degree of paired-associate learning in retardates. *Journal of Experimental Psychology*, 1971, **88**, 196–204.

BIRNBAUM, I. M. Context stimuli in verbal learning and the persistence of associative factors. *Journal of Experimental Psychology*, 1966, **71**, 483–487.

---

[5] Our T:R availability estimates are compared only with estimates from other word-trigram experiments because this is the only type of redundant-cue study in which the positional orientation of the cues during PA learning has been allowed to vary across studies (e.g., left–right vs. above–below configurations); the importance of positional orientation is suggested in the present discussion.

COHEN, J. C., & MUSGRAVE, B. S. Effect of meaningfulness on cue selection in verbal paired-associate learning. *Journal of Experimental Psychology*, 1964, **68**, 284–291.

HARRINGTON, A. L. Effects of component emphasis on stimulus selection in paired-associate learning. *Journal of Experimental Psychology*, 1969, **79**, 412–418.

HOUSTON, J. P. Stimulus selection as influenced by degrees of learning, attention, prior associations, and experience with the stimulus components. *Journal of Experimental Psychology*, 1967, **73**, 509–516.

JAMES, C. T., & GREENO, J. G. Stimulus selection at different stages of paired-associate learning. *Journal of Experimental Psychology*, 1967, **74**, 75–83.

KRUEGER, W. C. F. The relative difficulty of nonsense syllables. *Journal of Experimental Psychology*, 1934, **17**, 145–153.

LOVELACE, E. A. Cue selection in paired-associate learning: Meaningfulness and position. *Psychonomic Science*, 1968, **19**, 45–46.

NOBLE, C. E. Measurements of association value (a), rated associations (a') and scaled meaningfulness (m') for the 2100 CVC combinations of the English alphabet. *Psychological Reports*, 1961, **8**, 487–521.

SPEAR, N. E., EKSTRAND, B. R., & UNDERWOOD, B. J. Association by contiguity. *Journal of Experimental Psychology*, 1964, **67**, 151–161.

STEINER, T. E., & SOBEL, R. Intercomponent association formation during paired-associate training with compound stimuli. *Journal of Experimental Psychology*, 1968, **77**, 275–280.

UNDERWOOD, B. J., HAM, M., & EKSTRAND, B. Cue selection in paired-associate learning. *Journal of Experimental Psychology*, 1962, **64**, 405–409.

WITMER, L. R. The association value of three-place consonant syllables. *Journal of Genetic Psychology*, 1935, **47**, 337–360.

(Received December 27, 1971)

ERRATUM

In the title of the article by W. Kintsch and D. Monk, page 25 of the June issue, the word inferences should be substituted for interferences so that it reads as follows:
"Storage of Complex Information in Memory: Some Implications of the Speed with which Inferences Can Be Made"

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 195–201

STORAGE ASPECTS OF NOUNS PRESENTED UNDER IMAGERY AND ACOUSTIC CODING INSTRUCTIONS[1]

LOWELL D. GRONINGER,[2] BRUCE BELL, WILLIAM CYMER, AND BERNARD WESS, JR.
University of Maryland, Baltimore County

The effects of imagery and acoustic coding instructions on the storage dimensions of nouns with object referents and on the discriminability of the resultant memory traces were tested by presenting 80 words to *S*s under different coding instructions followed by a recognition memory task where the words were paired with distractors, each pair having similar images, similar sounds, or no similarities. Instructions to use imagery increased the number of word pairs correctly discriminated and decreased errors for all distractor types. There was no significant interaction between the type of instructions and the type of distractor. The results were interpreted as showing that: (*a*) instructions to use imagery increase the discriminability of memory traces for nouns with object referents, (*b*) there is no evidence for a qualitative change in storage dimensions with instructions to concentrate on either an image or a sound component of nouns with object referents.

Paivio (1969) and Bower (in press) have documented that imagery as a component of a word and instructions to use imagery are powerful factors in recall using relational paradigms (e.g., paired-associate learning), where images are presumably used as mediators. However, little attention has been given to the processes involved in the storage of individual items under imagery instructions. If one adopts the view that the storage of items is on a multidimensional basis (Anisfeld & Knapp, 1968; Wickens, 1970), an imagery dimension could be dominant, particularly for words that are easy to image such as words that refer to objects. The results of Rowe and Paivio (1971a) provide some evidence for this view by showing that individual words, when presented with imagery instructions, are more discriminable in a verbal discrimination task than individual words that were rehearsed by having *S*s repeat them aloud.

The purpose of the present study was to investigate further the nature of an imagery dimension as it relates to the storage of individual, object-referent words. This general problem can be subdivided into the following questions. First, since verbal labels are often associated with images, as occurs when words are used as stimuli, one major empirical question involves how these dimensions interrelate in the storage of image-evoking words. A second question concerns the extent to which these dimensions can be manipulated by different attentional sets initiated by *S*s. A third question, which is dependent on the second, involves further investigation of the findings of Rowe and Paivio (1971a) concerning the effectiveness on trace discriminability resulting from imagery and other types of instructions.

Paivio and Csapo (1969) have shown that the separation of an image from its verbal label can be accomplished with rapid presentation of materials. The interest of the present study, however, was in determining if a separation of verbal labels with images could occur with different attentional sets of *S*s. Findings from this manipulation should be of value in determining the extent to which an image code, independent of a verbal code, can be initiated and utilized by an individual at his discretion. If it is possible, through instructions to *S*s, to separate an image from its verbal label, different types of errors should be apparent on a recognition memory task based on the similarity of the distractor item to the major dimension of coding, i.e., imagery or sound. With imagery instructions, there should be more errors and less items correct, relative to controls, when distractors are similar on an image dimension, while a similar occurrence would be predicted when the distractors are homonyms, with instructions to concentrate on the sound of the words. The latter prediction is supported by the results of Cermak, Schnorr, Buschke, and Atkinson (1970). Another possibility would be that the imagery dimension might be dominant regardless of the type of instructions, in which case the distractors that are similar in imagery to the correct items would decrease recall and cause the most errors for all conditions.

---

[1] This research was conducted by the junior authors under the direction of the senior author as a class project in an advanced experimental psychology seminar.

[2] Requests for reprints should be sent to Lowell D. Groninger, Department of Psychology, University of Maryland, Baltimore County, Baltimore, Maryland 21228.

METHOD

Subjects.—The *S*s were introductory psychology students who had signed to participate in 4 hr. of experiments as a substitute for other course work. Seventy-two *S*s contributed the basic data of the experiment while 59 additional *S*s were used in evaluating the lists themselves.

Experimental design and materials.—The experiment had a 3 × 3 mixed design with repeated measures on the second factor. The first factor was instructions (I). The three types of instructions representing conditions were Imagery, Control, and Acoustic. In the Imagery condition, *S*s were told to concentrate as much as possible only on the image of the word. In the Control condition, *S*s were told simply to learn as many of the words as they could, while *S*s in the Acoustic condition were told to try to concentrate as much as possible only on the sound of the word, while repeating the word aloud, without thinking of its image. All *S*s were told that they would later be asked to identify the words on a recognition memory task. The second factor was distractors (D), representing the type of distractor used in the recognition memory task. The distractors were similar either on an Imagery dimension (nouns with different labels having a common referent), an Acoustic dimension (homonyms), or had no apparent similarities (Control). The *S*s were nested in groups of 24 within the instructions factor.

All words used in the experiment were nouns that were judged by *E*s to have object referents. The assumption that was made in using words with object referents was that these words should be capable of evoking a "first-order" image from *S*s. For example, "book" would likely produce a direct image related to the concept of a book whereas "democracy" might produce an image of a legislative building, a group of people voting, or some other image associated with the meaning of democracy. It also seems reasonable to assume that these words would score high on an imagery scale such as used by Paivio, Yuille, and Madigan (1968). To test this assumption, the words were rated by 34 *S*s along with 14 other words chosen from the norms of Paivio et al. to represent intervals throughout the 7-point scale. The instructions of Walker (1970), including his slight modification of the Paivio et al. procedure, were used. Three of the first 10 words rated by *S*s were words from the lower half of the Paivio et al. norms to assist *S*s in forming representative anchors. The mean ratings for the Similar Image, Neutral, and Homonym word groups were 5.81, 5.59, and 5.43, respectively. The correlation between the 14 common words from the present ratings and those of Paivio et al. was .97. A comparison of mean ratings of the 14 common words showed a mean difference of .26 with the ratings of this study being lower. It seems fair to say that the criteria used in selecting the current words overlap considerably with the criteria that *S*s use in judging words to be high on an image-evoking scale.

The words used in the study were chosen without regard to word frequency, a dimension which does not seem to be important in verbal discrimination tasks when the words are also high in imagery ratings (Rowe & Paivio, 1971b). The list of 60 word pairs shown in Table 1 were constructed to be used alternately between *S*s as learning items and distractors. Twenty word pairs were homonyms, 20 had no apparent similarities, and 20 were words that the four *E*s rated as creating a similar image with different labels. The pairs of homonyms used in the study were verified as giving different referents by 25 different *S*s who were shown the homonyms in scattered fashion for 1 sec. on a screen and given 7 sec. to write an association or function of the word. The criterion for inclusion in the list was that at least 75% of these *S*s had to give responses that were judged to indicate different referents for each homonym in a pair. The judging was by consensus of the four *E*s. Of the homonym pairs used in the experiment, the average rating of different referents was 90%. This control was necessary to assure that

TABLE 1

LIST OF TEST-DISTRACTOR PAIRS

Similar image	Neutral	Homophone
Box-Carton	Vaccine-Nickel	Vein-Vane
Soil-Dirt	Chick-Broom	Beer-Bier
Stream-Creek	Man-Ink	Night-Knight
Shrub-Bush	Dress-Bay	Isle-Aisle
Street-Road	Root-Lava	Dough-Doe
Rug-Carpet	Whale-Tent	Oar-Ore
Officer-Policeman	Guest-Cog	Surf-Serf
Trousers-Pants	Plate-Dot	Air-Heir
Auto-Car	Desk-Peach	Hare-Hair
Infant-Baby	Fog-Tube	Stake-Steak
Wharf-Pier	Crock-Reef	Cash-Cache
Bag-Sack	Snow-Zoo	Tee-Tea
Chimney-Smokestack	Acid-Cable	Yolk-Yoke
Pamphlet-Leaflet	Eye-Hay	Arc-Ark
House-Home	Gadget-Fender	Waist-Waste
Mud-Mire	Solvent-Theater	Sail-Sale
Treasure-Riches	Mallet-Kidney	Ant-Aunt
Froth-Foam	Candle-Asphalt	Ferry-Fairy
Rock-Stone	Book-Salt	Kernel-Colonel
Branch-Limb	Nut-Lamp	Mail-Male

if *S*s were using imagery, they would have a clear choice between homonym pairs in the recognition memory task with a minimum influence from vocabulary or spelling deficiencies on the part of *S*s. In addition to the 60 words used in the experimental conditions, 20 buffers of the same word type were presented to *S*s, but not tested, when pilot data indicated a potential ceiling effect. The buffers occupied the first five positions of the list and were randomly placed throughout the rest of the list.

Procedure.—The *S*s participated in the experiment in groups of four or less. Groups were assigned to conditions in alternating fashion. The specific instructions to *S*s for the Imagery instruction group were:

> This is an experiment in imagery. I am going to show you 80 words, one at a time, at a 5-second rate. The words will appear on the screen for 1 second followed by a 4-second blank time followed by another word, etc. As soon as you see a word please close your eyes and try to form a distinct mental image of the object represented by that word, and concentrate only on that object until the next word is shown. Please try not to repeat the sound of the word at the same time that you are picturing the word. This is difficult to do but essential for the success of our experiment. For example, if the word is "lion," think only of the image of a lion without at the same time repeating the sound of the word to yourself. After the words have been shown you will be asked to identify some of the images in a recognition task. You will first have 7 practice words. Any questions?

The specific instructions for the Acoustic instruction condition were:

> This is an experiment on memorizing sounds. However, instead of memorizing something like a tune, I am going to ask you to remember sounds of words. I will show you 80 words one at a time, at a 5-second rate on the screen. You will see each word for 1 second followed by a 4-second blank time followed by another word, etc. As soon as you see a word say it aloud and listen carefully to the sound components. Keep repeating the word aloud, in time to the beat you will hear through your headphones. Try to concentrate only on the sound components and rhythm of the word without thinking of the image of the word as you repeat the word aloud. This is difficult to do, but essential to the success of our experiment. For example, if the word is "lion," concentrate only on the sound of each phoneme, the *la* sound, the *i* sound, the *ah* sound and the *un* sound—la i ah un, la i ah un, la i ah un. After the words have been shown you will be asked to identify some of the sounds in a recognition task. You will first have 7 practice words. Any questions?

The specific instructions for the Control condition were:

> In this experiment, I am going to present 80 words to you, one at a time, at a 5-second rate. A word will appear on the screen for 1 second followed by a 4-second blank time followed by another word, etc. Your task is simply to try to learn as many items as you can. After the words have been shown you will be asked to identify some of the words in a recognition task. You will first receive 7 practice words. Any questions?

After the words were presented, in random order with respect to the conditions that they represented, *S*s were given a sheet of paper containing an alternating arrangement, with respect to conditions, of the 60 paired words that were to be tested. The *S*s were told to check the member of each pair that they felt had just been presented to them and state their confidence in their decision. The confidence scale consisted of three numbers where "1" was very confident, "2" was think so but could be wrong, and "3" was just guessing. The *S*s were then given a questionnaire asking them to make an estimate of the average time they spent during the presentation of the words on the following: thinking of the image of the word; thinking of the sound of the word; thinking of prior words; other. In addition, *S*s were asked if and how many words they might have missed seeing. Data from *S*s were discarded if their questionnaires showed that they had not spent the largest amount of time during the presentation of the words concentrating on the sound of the word if they were in the Acoustic condition, or concentrating on the image of the word if they were in the Imagery condition. Data from *S*s were also discarded for those *S*s who reported that they missed seeing more than one item. Each group had an average reported miss rate of .004 per word. Fifteen *S*s were discarded for failure or inability to follow instructions according to the previously mentioned criteria regarding sound and imagery, and 7 were discarded for missing more than one item. In discarding data, the questionnaires were kept separate from the remainder of *S*s' data to assure that rejection decisions, based on questionnaires, would be independent of all other data. The *S*s that were eliminated on the basis of questionnaires were replaced by additional *S*s in order to maintain an *N* of 24 in each cell.

The words were shown in a darkened room with the projector beam blocked between presentations. The purpose of the darkened room was to facilitate concentration and to provide a contrast background from which the flash of the words on the screen would attract the attention of *S*. An additional attentional cue was the sound of the projector changing slides. The 1-sec. showing of the word was accomplished manually with a projection tachistoscope set at a 1-sec. exposure. The purpose of the 1-sec. showing followed by the 4-sec. blank time was to minimize the effects of perceptual cues involved in the physical structure of the words.

In the Acoustic condition, *S*s rehearsed the words out loud in time to a 1-sec. beat of the metronome. The *S*s in this condition were required to wear headphones to block out the sound of others rehearsing. The sound of the metronome was amplified over a speaker system so it could be heard over the muf-

TABLE 2

AVERAGE PERCENT OF TIME THAT SUBJECTS REPORTED ATTENDING TO VARIOUS COMPONENTS OF THE TASK

Instruction cond.	Word sound	Word image	Prior words	Other
Imagery	17.04	66.67	9.79	6.50
Control	23.13	28.33	29.38	19.17
Acoustic	56.13	17.50	13.04	13.34

fling effect of the headphones. The *S*s were also placed about 12 ft. apart during this condition.

RESULTS

The average amount of time that *S*s in each condition reported spending thinking of the image or sound of a word, thinking of prior words, or thinking of something else is shown in Table 2. From Table 2 it is apparent that the instructions did differentiate attention times between imagery and sound. However, in no case did an *S* report total attention to one category.

The data were scored according to three criteria: (*a*) number correct and number of misses for items with a confidence rating of 1; (*b*) number correct and number of misses for items with a confidence rating 1 or 2; (*c*) number correct and number of misses for all data. The purpose of separating data based on confidence ratings was to increase the sensitivity of the design. The highest confidence rating, 1, contains very few errors as shown in Table 3, and should most accurately reflect the effects of the conditions. The percentage of number of items correct over the total number of items marked within each confidence rating is shown in Table 4. The scores in Table 4 demonstrate that the confidence ratings were separating data along a sensitivity dimension. The only exception to this conclusion, ratings of 3 in the Imagery

TABLE 3

MEAN NUMBER CORRECT AND ERRORS WITH CONFIDENCE RATINGS OF 1, OF 1 OR 2, OR USING ALL DATA FOR THE TYPES OF DISTRACTORS AS A FUNCTION OF INSTRUCTIONS

Instruction cond.	Distractor type							
	Similar image		Neutral		Homophone		$\bar{X}$	
	Correct	Errors	Correct	Errors	Correct	Errors	Correct	Errors
				"1's"				
Imagery	16.13	.21	18.25	.08	17.83	.63	17.40	.31
Control	12.54	.75	14.80	.46	14.66	.58	13.33	.60
Acoustic	13.29	.38	14.08	.17	14.66	.42	14.01	.32
$\bar{X}$	13.97	.45	15.44	.24	15.72	.54	14.91	.41
				"1's" or "2's"				
Imagery	18.50	.79	19.25	.29	18.54	1.08	18.76	.72
Control	15.50	1.79	16.00	1.33	16.83	1.25	16.11	1.46
Acoustic	16.58	1.29	17.00	.70	16.63	1.46	16.74	1.15
$\bar{X}$	16.86	1.29	17.42	.77	17.33	1.26	17.20	1.11
				All data				
Imagery	18.87	1.13	19.58	.42	18.95	1.05	19.14	.86
Control	16.92	3.08	17.54	2.46	17.91	2.09	17.46	2.54
Acoustic	17.79	2.21	18.41	1.59	17.79	2.21	18.01	1.99
$\bar{X}$	17.86	2.14	18.51	1.49	18.24	1.76	18.20	1.80

Note.—Based on 20 item lists.

TABLE 4

Percentage Correct within Confidence Ratings of 1, 2, or 3

Instruction cond.	Distractor type								
	Similar image			Neutral			Homophone		
	1	2	3	1	2	3	1	2	3
Imagery	387/392 .99	57/71 .80	9/17 .53	438/440 .99	24/29 .83	8/11 .73	428/443 .97	17/25 .68	10/12 .83
Control	301/329 .94	71/95 .75	35/66 .53	336/347 .97	48/68 .71	37/65 .57	352/366 .96	52/64 .81	30/50 .60
Acoustic	319/328 .97	79/97 .81	33/55 .60	338/342 .99	70/82 .85	34/56 .61	352/362 .97	47/72 .65	29/46 .63

Note.—Total *N* was 480, based on 20-item lists for 24 *S*s. Numbers 1, 2, and 3 indicate confidence ratings.

instructions condition, had an *N* too small to give a reliable percentage.

An analysis of variance for the data with a confidence rating of 1 showed there was a significant difference between instruction conditions with the Imagery group obviously superior to the other groups, F (2, 69) = 9.08, $p < .01$. The type of distractor was also significant, F (2, 138) = 19.36, $p < .01$, with the major source of variance coming from a decrement in the recall of the Similar Image words. The I × D interaction that would be predicted on the basis of different instructions creating coding on different dimensions was not found, F (4, 138) = 1.17, $p > .05$. In fact, the four cells where the interaction would be predicted (II, IA, AI, AA) show very similar patterns. The Imagery instruction condition had an average of 8.50% fewer items recalled when the distractor words were similar in image compared to when the distractors were homophones, while a value of 6.85% for this difference was obtained when *S*s were instructed to concentrate only on the sound of the words. The analysis of the data for the combined 1 and 2 ratings showed a significant effect for instructions, F (2, 69) = 7.72, $p < .01$, and a marginally significant effect for distractor type, F (2, 138) = 2.99, $p < .06$. The I × D interaction was marginally significant overall, F (4, 138) = 2.31, $p < .1$, but almost nonexistent in the four cells of major interest.

The analysis using all the data, irrespective of confidence ratings, showed a significant effect for instruction conditions, F (2, 69) = 7.62, $p < .01$, with Imagery instructions having the highest scores, but no significant difference ($\alpha = .05$) for the I × D interaction. Distractor type was significant, F (2, 138) = 5.28, $p < .01$, with the Similar Image group having the lowest scores. As can be seen in Table 3, the pattern of results in the number correct data is similar over the range of confidence ratings with the effects most pronounced at the "1" rating, where the sensitivity is greatest.

An analysis of errors made on the data having the highest confidence rating is shown in Table 3. Although, within any single confidence rating, these data are independent of the number correct data, the pattern of results is somewhat similar. The Imagery instruction condition and the Acoustic instruction condition have the fewest errors, although the difference between instruction conditions was not significant, F (2, 69) = 1.70, $p > .05$, perhaps due to a ceiling effect. The I × D interaction also was not significant. Type of distractor was significant, F (2, 138) = 3.81, $p < .05$, with the Neutral distractors having the fewest errors. For misses with a confidence rating of 1 or 2, the differences between instruction conditions were significant, F (2, 69) = 3.30, $p < .05$, with the Imagery group having the fewest errors and the Control group the most. Distractor type was also significant, $F(2, 138) = 5.29$, $p < .01$, with the Neutral group having the fewest errors. The I × D interaction was not significant. Because of the forced-choice technique used in the experiment, the error classification for all the data show the same effects as the number correct classification for all the data.

DISCUSSION

The most striking feature of the results is the magnitude of the effect of the Imagery instructions on word discrimination. Within confidence ratings of "1," Imagery instructions increased the number of pairs correctly discriminated in all distractor conditions by about 25%. Since the words were presented individually during learning rather than in conjunction with other words, as occurs in verbal discrimination paradigms (i.e., Rowe & Paivio, 1971a), this study provides evidence for Imagery instructions directly effecting the storage of object-referent words. With the present paradigm, the distractor item not occurring during the initial presentation, *S*s had no opportunity to store discriminating features of word pairs along with the words. Therefore, the present study gives additional confirmation to the finding of Rowe and Paivio (1971a) that high-imagery words can be made more discriminable with specific instructions to concentrate on the images of the words.

There does appear to be an interesting paradox in the results. Imagery instructions had a highly significant effect on the number of word pairs for which a correct discrimination was made, but there was no significant Type of Instruction × Type of Distractor interaction. It was expected that *S*s would not do as well on distractor items similar in image when given Imagery instructions as opposed to Control instructions. One partial solution to this paradox is to assume that the attentional sets induced by the instructions did not create single exclusive coding dimensions; instead the imagery set made the words more discriminable by strengthening an imagery component of the words without at the same time weakening the other dimensions which aid in word discrimination. Thus, word pairs similar in image could still be efficiently discriminated. This view can account for the relatively accurate performance of *S*s given Imagery instructions combined with Similar Image distractors, but it still does not account for the finding that Imagery instructions increased discrimination to a similar degree for all distractor types. A more complete explanation might be to note that while the referents in the Similar Image group were similar, they were, in many cases, probably not identical. Therefore, the effect of Imagery instructions may have been to increase the distinctiveness of these referent pairs, as well as the referent pairs of the other groups, thus increasing the number of word pairs correctly discriminated in all distractor groups.

The effect of distractor type on discrimination of the word pairs also merits comment. The Similar Image group had poorest discrimination under all instruction conditions compared to the Control group. This decrement is not surprising in view of other recognition memory studies (Anisfeld & Knapp, 1968; Buschke & Lenon, 1969; Cermak et al., 1970) that have found confusion resulting from words that were semantically similar. Unfortunately, words with similar referents but different labels have the property of being semantically similar as well as similar on an image dimension. Whether increased confusion in the Similar Image group results from a confusion of images or from semantic confusion of some type cannot be determined from this study. The results from Table 1 demonstrate that images were present in all conditions, making the imagery explanation at least tenable. In any case, the discrimination or decoding problem was not extremely severe since even though discrimination was relatively lower in the Similar Image group, the total correct was still high.

The present study did not confirm the decrement in discriminating homonym word pairs as found by Cermak et al. (1970) and Buschke and Lenon (1969). However, these studies used words of all types rather than object-referent nouns as used in this study. The implication is that when words have distinct object referents, these are sufficient to enable word-pair discrimination, independent of the similarity of sounds involved in the word labels.

As previously pointed out, the results of this study indicate that Imagery instructions can make a memory trace more distinct when items are words with object referents. Thus it appears that instructions to use imagery not only facilitate the relational aspects of words as documented earlier, but, contrary to the findings of Bower (1970), Imagery instructions can, under the proper conditions, affect the storage of items as reflected in a recognition memory task. Based on the results of Rowe and Paivio (1971a) and Bower (1970), these conditions for imagery effectiveness in recognition memory involve the formation of single images as opposed to compound images. Further research is needed to provide an adequate explanation of why single images behave differently than compound images within recognition memory.

REFERENCES

Anisfeld, M., & Knapp, M. Association, synonymity, and directionality in false recognition. *Journal of Experimental Psychology*, 1968, **77**, 171–179.

Bower, G. H. Imagery as a relational organizer in associative learning. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 529–533.

Bower, G. H. Mental imagery and associative learning. In L. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, in press.

Buschke, H., & Lenon, R. Encoding homophones and synonyms for verbal discrimination and recognition. *Psychonomic Science*, 1969, **14**, 269–270.

Cermak, G., Schnorr, J., Buschke, H., & Atkinson, R. C. Recognition memory as influenced by differential attention to semantic and acoustic properties of words. *Psychonomic Science*, 1970, **19**, 79–81.

Paivio, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, **76**, 241–263.

Paivio, A., & Csapo, K. Concrete image and verbal memory codes. *Journal of Experimental Psychology*, 1969, **80**, 279–285.

Paivio, A., & Rowe, E. J. Noun imagery, frequency, and meaningfulness in verbal discrimination. *Journal of Experimental Psychology*, 1970, **85**, 264–269.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 1968, **76**(1, Pt. 2).

Rowe, E. J., & Paivio, A. Imagery and retention instructions in verbal discrimination and incidental paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 668–672. (a)

Rowe, E. J., & Paivio, A. Word frequency and imagery effects in verbal discrimination learning. *Journal of Experimental Psychology*, 1971, **88**, 319–326. (b)

Walker, H. J. Imagery ratings for 338 nouns. *Behavioral Research Methods and Instrumentation*, 1970, **2**, 165–167.

Wickens, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–14.

(Received December 29, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 202–210

SEMANTIC GENERALIZATION OVER A BIPOLAR DIMENSION OF MEANING[1]

PETER A. ORNSTEIN[2]

Princeton University

AND

DAVID A. GRANT, WILLIAM C. WATTERS

University of Wisconsin

In two experiments, semantic generalization was tested along a bipolar dimension of connotative meaning. During acquisition, *S*s were presented with Good stimulus words, i.e., words with positive connotations, Bad stimulus words, i.e., words with negative connotations, or with both types of material. The task of the *S*s was to press a button so as to anticipate the illumination of one of two small lights which was contingent on the connotation of the stimulus. During generalization testing, all *S*s were presented with 15 stimuli from various points on the Good–Bad continuum. Significant generalization gradients were obtained, with *S*s responding most frequently to stimuli whose connotations were similar to the training words, less to neutral stimuli, and still less to those stimuli which were most dissimilar in connotation. The gradient of semantic generalization following differential conditioning was only slightly steeper than that following single-cue training. Complex assimilation and context effects among the verbal stimuli appear to shift the "effective" position of the words from their rated positions on the Good–Bad continuum, so that precise specification of the shape of the semantic generalization gradient may prove to be difficult.

While semantic generalization is a well-documented, if somewhat elusive phenomenon (cf. Feather, 1965; Maltzman, 1968; Razran, 1961), investigators working with classical conditioning techniques have had considerable difficulty in demonstrating the existence of stable monotonic gradients of generalization (e.g., Cole & Williams, 1966; Lang, Geer, & Hnatiow, 1963; Maltzman & Langdon, 1969; Mednick, 1957; Perry, Ornstein, & Grant, 1971). In contrast, semantic generalization is more regularly obtained with instrumental conditioning. In these situations (e.g., Dicken, 1961; Mink, 1963), *S*s are trained to respond to a set of words by pressing a button or pushing a lever, and generalization is typically tested with a second list of words containing the original words plus additional related test and unrelated control words. Kurcz (1964) modified this technique to permit recording of the electromyogram (EMG) from the muscles active in key pressing, in addition to overt responses. Using this procedure, Cramer (1970c) has obtained the most convincing evidence of a gradient of semantic generalization yet presented. Generalization was shown to be a linear function of the associative strength between the training stimuli and the test stimuli.

The major purpose of the present experiments was to use an instrumental response to study the form of semantic generalization gradients obtained following single-cue and discrimination training procedures. The literature on primary stimulus generalization indicates that there is less generalization to distant test stimuli following differential training than following single-cue training (e.g., Hoving, 1963; Jenkins & Harrison, 1960; Perkins, Hershberger & Weyant, 1959; Reingold & Perkins, 1955). While these differences in the slopes of gradients have been of some theoretical interest (e.g., Hull, 1943; Lashley & Wade, 1946; Spence, 1936), the problem has not been studied in the context of semantic generalization. The only experiment bearing on this point (Novak & Grant, 1965) yielded some evidence for less generalization following differential conditioning, but did not examine the shapes of gradients, and is difficult to interpret because of the relative lack of generalization observed.

---

[1] This work was supported in part by United States Public Health Service Grant MH 06792 to D. A. Grant, United States Public Health Service predoctoral fellowship to P. A. Ornstein, and United States Public Health Service Grant MH 10742 to S. Glucksberg. This report was written while the first author held a North Atlantic Treaty Organization Postdoctoral Fellowship at University College, London. The authors express their thanks to Barbara Witten for running *S*s in Exp. II, and to Irving Maltzman for helpful comments on an earlier version of this paper.

[2] Requests for reprints should be sent to Peter A. Ornstein, Department of Psychology, Green Hall, Princeton University, Princeton, New Jersey 08540.

A secondary purpose of these experiments was that of establishing a reliable procedure for obtaining semantic generalization, because when generalization has been obtained the relative amounts of observed semantic and phonetographic generalization (Hartman, 1965; Razran, 1961) are influenced by many important variables. These variables include age (Reiss, 1946; Rice & DiVesta, 1965), administration of drugs (Shvarts, 1960), or alcohol (Levy & Murphy, 1966), and presence of behavior pathology (Peastrel, 1964). Development of a technique which could be used to obtain reliable semantic generalization in a base-level sense is needed to facilitate further investigation of the phenomenon and the factors influencing it.

The procedure adapted here is somewhat similar to that recently used by Cramer (1970a, 1970b, 1970c), whose work was unavailable when these experiments were initiated, but it also resembles in part that used by Buss (1961) with an instrumental response, and by Lang et al. (1963) with classical conditioning. An instrumental button-pushing response was chosen, and, instead of recording implicit muscle responses, the force exerted by *S*s in depressing the buttons was monitored. In contrast to Cramer's experiments, training stimuli were not used in generalization testing, so that gradients of generalization uncontaminated by further discrimination training during testing could be obtained. The *S*s were trained to respond to a set of words sharing a common meaning component, as defined by the Good–Bad scale of the semantic differential (Osgood, Suci, & Tannenbaum, 1957), which was employed because a well-defined bipolar dimension is necessary if the shape of the gradient of generalization is to be investigated. The polar extremes of this dimension are, in fact, opposites, and an intermediate point exists within the "neutral" category. Such a dimension was used by Acker and Edwards (1964) in a classical conditioning study, following up work of the Soviet investigator Volkova (1953, 1957). In the Acker and Edwards experiment, however, *S*s were trained to discriminate between the words "Good" and "Bad," and were then tested for generalization to words whose connotations varied on the Good–Bad continuum.

EXPERIMENT I

Method

Experimental design.—Four major groups of 20 *S*s each were formed by the factorial combination of two variables, training and evaluation. Training refers to whether the acquisition stage involved single-cue or differential conditioning. Evaluation refers to whether *S*s were instructed or not instructed to evaluate the connotative meaning of each stimulus word. Each of these four groups was subdivided so as to counterbalance the end of the Good–Bad continuum which was positively reinforced during training. These Good and Bad subgroups formed a positive-stimulus variable. All *S*s received generalization trials with words from three levels of a word-class variable, Good, Neutral, and Bad words.

Stimulus materials.—The stimuli were selected from the semantic atlas of Jenkins, Russell, and Suci (1958). Good stimuli had average ratings of 1.0 to 2.0 on the Good–Bad scale; Neutral stimuli had average ratings of 3.5 to 4.5; and Bad stimuli had average ratings of 6.0 to 7.0, in the Jenkins et al. atlas. Within each group of stimulus words, frequency, as determined by the Thorndike-Lorge count, and part of speech, were balanced as well as possible. The Good words used during training were as follows: FLOWER, MUSIC, CLEAN, RADIANT, HARMONIOUS, while the Bad words were the following: HATE, DEVIL, HURT, PUTRID, DISCOMFORT. The generalization stimuli were as follows: Good—MOTHER, SWEET, HEAL, GRACEFUL, SUNLIGHT; Neutral—HEAVY, BLOCK, RIGID, QUOTA, MOLD; Bad—SIN, RAGE, FRAUD, STAGNANT, CRIMINAL.

Apparatus.—The words were projected by a Kodak Carousel slide projector onto the back of a rectangular 20.3 × 25.4 cm. ground-glass screen which was approximately 60 cm. in front of S's eyes. The slides appeared as white letters on a dark background, and the projected letter height was approximately 2.54 cm. The S was seated in a modified desk chair with two buttons in the center. These buttons were attached to two Grass Instrument force displacement transducers (Model FT .03C), and response pressure was recorded by an ink-writing two-channel Texas Instrument Oscillo/Riter. A complete depression of the response button caused a pen deflection of 20 mm. In front of S was a small gray box upon which were mounted two small lights, which corresponded to the two buttons on the chair. Equipment sounds were masked by approximately 80 db. of white noise delivered through earphones.

Procedure.—All Ss were told that they would see a series of words, each followed often by the illumination of one of the two small lights in front of them. The task of the Ss was to learn to associate the words and the corresponding light, and to anticipate the onset of the appropriate light by depressing the corresponding button when the word appeared. They were told that they must respond on each trial, but that on some trials no lamp would be illuminated. Further, one-half of the Ss in each group, as defined by the evaluation variable, were told to evaluate subvocally the connotation of each word as it was presented, while the other half of the Ss were not so instructed.

The S's arm rested on the arm of the chair, with her index finger on the left button and her middle finger on the right button. During initial training, Ss in the simple acquisition groups were presented with 15 trials of either Good or Bad words. The five training words were each presented three times. In these groups, the presentation of the stimulus word was followed always by the illumination of the right-hand light. The Ss in the discrimination training groups were presented with 30 trials of Good and Bad stimuli randomly intermixed. The five Good and five Bad stimuli were randomized within each of three 10-trial blocks. The Good stimuli were followed always by illumination of the right-hand light and the Bad stimuli were followed by illumination of the left light in one subgroup, and this procedure was reversed in a second subgroup, as defined by the positive-stimulus variable. The generalization test, given all Ss, consisted of the presentation of 15 generalization words, each presented once. The reinforcement lights were not used during the generalization test. Two different word orders were employed during generalization testing.

Each slide was presented for 1,200 msec. During initial training, following an interstimulus interval of 800 msec., one of the two lights was illuminated for 400 msec. The intertrial intervals ranged from 5 to 15 sec., averaging 10 sec., and were programmed by a Western Union tape transmitter.

Subjects.—A total of 80 right-handed women students, enrolled in introductory psychology courses at the University of Wisconsin, were randomly assigned to the eight experimental groups. All Ss received class credit for their participation. Ten additional Ss were discarded, 8 because of equipment failure, 1 because she reportedly did not understand the meaning of the words, and 1 because she did not respond at all.

Results

The response measure employed was frequency of right-hand button depressions. To be scored, a response had to produce at least a 1-mm. pen deflection following a 200-msec. level base line. Response made during the first 800 msec. of stimuli duration during training and 1,200 msec. of stimulus duration during generalization testing were scored.

The simple anticipatory responses were learned equally well by all Ss during initial training. An analysis of variance, performed on the number of responses made to the positive stimuli during the last five positive acquisition trials, indicated no differences among the groups. Further, the differential conditioning groups, which were exposed to both Good and Bad stimuli during acquisition, all showed good differential conditioning. An analysis of variance, performed on terminal discrimination data, indicated no differences between treatments; only the difference between frequency of responses to positive and negative stimuli was significant, $F(1, 36) = 657.79, p < .001$.

The generalization data, in the form of mean right-button responses as a function of class of generalization stimulus words are presented, for each of the eight groups, in the upper panel of Fig. 1. Despite differences in the form of the gradient among the various groups, semantic generalization along a bipolar meaning dimension appears to have been demonstrated. The Ss trained to respond to stimuli at one extreme of a meaningful dimension responded most to stimuli of similar connotative meaning, less to neutrally toned stimuli, and still less to those words which were polar opposites. An orthogonal polynomial analysis of variance (Grant, 1956) based on the number of right-hand responses made to each of the three categories of generalization stimuli was performed. In this analysis, the word-class

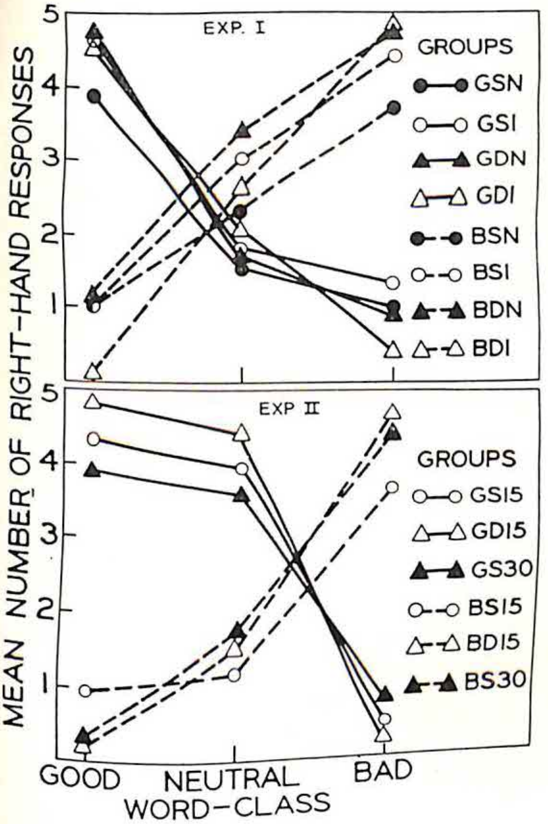


FIG. 1. Mean number of right-hand responses as a function of type of test stimuli for all groups in Exp. I and II. The *S*s received either single-cue (S) or differential (D) training with either Good (G) or Bad (B) words being reinforced. The *S*s were instructed (I) or were not instructed (N) to evaluate covertly the connotations of the words in Exp. I. There were 15 or 30 trials with positive stimuli during acquisition in Exp. II.

variable (Good, Neutral, Bad) was rearranged into a semantic-distance-from-training-stimulus variable. Thus, the responses to the Bad stimuli for the Bad groups, and those to the Good stimuli for the Good groups were combined into a "near" category, etc. The only significant main effect was due to semantic distance, F (2, 144) = 221.53, $p < .001$, $MS_e =$ 1.16, reflecting the observed generalization effect. This semantic-distance effect was composed of highly significant linear, F (1, 72) = 365.22, $p < .001$, $MS_e = 1.39$, and quadratic, F (1, 72) = 6.77, $p < .05$, $MS_e = .93$, components. Further, subsequent two-tailed t tests indicated that all three points differed significantly from each other.

There were three significant interactions. Inspection of Fig. 1 indicates that more responses were made to the Neutral stimuli by those *S*s initially trained to respond to the Bad words than by those trained to respond to the Good stimuli. This resulted in a significant Positive Stimulus × Distance interaction, F (2, 144) = 6.85, $p <$ .005, $MS_e = 1.16$. This interaction is due largely to very significant differences between the quadratic trend components of the two positive-stimulus groups, F (1, 72) = 16.96, $p < .001$, $MS_e = .93$, reflecting the greater curvature of the gradient of the Good group. Subsequent two-tailed t tests indicated that the Bad groups made significantly more responses to the Neutral stimuli than did the Good groups, t (216) = 4.06, $p < .01$, but that the groups did not differ at the near and far semantic distances.

The second significant interaction, that of Training × Evaluation, F (1, 72) = 5.38 $p < .05$, $MS_e = 1.71$, can be seen in Fig. 1. The discrimination training groups made more responses than did the *S*s given single-cue training under conditions of no instructions, while the reverse was true of the instructed *S*s. Subsequent two-tailed t tests revealed that the only significant differences were between uninstructed *S*s given discrimination training and uninstructed *S*s given single-cue training, t (72) = 2.17, $p < .05$. The corresponding difference under instructed conditions did not approach significance.

The final significant interaction, Training × Distance, F (2, 144) = 4.94, $p < .01$, $MS_e = 1.16$, is the most important for the purposes of this investigation. An examination of the shape of the generalization gradients following single-cue conditioning and differential conditioning shows that sharper or steeper gradients of generalization were obtained following discrimination training. Confirming visual examination of these curves, the trend analysis indicated that the Training × Distance interaction was largely a function of differences between the two training groups in linear trend components, $F\ (1, 72) = 7.55$, $p < .01$, $MS_e = 1.39$. Subsequent two-tailed t tests indicated that the discrimination training groups were significantly higher at the near point, $t\ (216) = 2.13$, $p < .05$, and marginally lower at the far point, $t\ (216) = 1.84$, $p < .10$.

Discussion

Semantic generalization gradients based upon the evaluation connotation of the verbal stimuli were apparently obtained following both single-cue and differential instrumental training. The major aim of the experiment was to compare the generalization gradients obtained following these two types of training procedures. As was predicted from the work on primary stimulus generalization (e.g., Jenkins & Harrison, 1960; Hoving, 1963), the gradient of generalization following discrimination training was sharper or steeper than that following single-cue reinforcement. This difference was reflected in the significant Training × Distance interaction, and more specifically, the significant differences between the groups in linear trend. It should be noted, however, that despite the significance of differential reinforcement effect, its magnitude seems not to be as great as has been found in some studies of primary stimulus generalization (e.g., Jenkins & Harrison, 1960).

The two other obtained interactions were unexpected. The significant Positive Stimulus × Distance interaction indicated that with the present stimuli at least, the transfer phenomenon was not symmetric. The interaction reflected the greater number of responses made to Neutral stimuli following training with Bad stimuli than with Good stimuli. This may indicate that "neutrality" with respect to polar extremes in the conditioning situation may be quite different from "neutrality" in rating words on the evaluative dimension.

The evaluation variable was included in the present experiment to determine if additional instructions to attend to the connotations of the words would influence the amount of observed semantic generalization. As was the case with Acker and Edwards' (1964) classical conditioning study, no effect of evaluation was found in the present experiment. The Training × Evaluation interaction, however, was significant and is difficult to interpret. One possibility is that the evaluation of connotative meaning is implicitly involved in the process of discrimination. It is reasonable to assume that *S*s in the discrimination groups were more likely to evaluate the words specifically in the absence of instructions to do so, so that the difference between the discrimination and the single-cue *S*s would not occur under the conditions of instruction where the response of evaluation was explicitly demanded of all *S*s.

Finally, although these data have been interpreted as providing an intermediate point in a semantic generalization gradient over a bipolar dimension, it should be noted that alternative interpretations are possible. First, this experiment may be viewed as a type of concept identification study. The Good and Bad words, while occupying polar positions on the Good–Bad scale, also form a "Good" (or "Bad") concept or class, and can readily be labeled as such by all *S*s. The Neutral words do not, however, form such an identifiable class. It is possible that the intermediate position of the responses to the Neutral words, which here has been interpreted as occurring because of semantic generalization, may arise because *S*s experienced some confusion attempting to classify the Neutral words into the "Good" or "Bad" categories of the training words. If the Neutral words also belonged to a readily identifiable category, *S*s might have less difficulty in distinguishing between them and the polar training words. Under these conditions, the apparent semantic generalization gradient might vanish, with responding being equivalent for the Neutral and nonreinforced (opposite) Good or Bad words. Second, with respect to the slope differences between the single-cue and differential conditioning conditions, it should be noted that the single-cue *S*s received 15 trials during training, while the discrimination *S*s received 30 trials, 15 with each of two kinds of words. Possibly the greater number of trials, regardless of type, could account for the steeper slope.

Experiment II was designed to explore these considerations. To determine if the obtained semantic generalization effect might be attributable to confusion in classifying the Neutral words, the Neutral words which were employed in the second experiment belonged to a clearly identifiable taxonomic category. Also, to explore further the slope differences between single-cue and differential conditioning, an additional single-cue acquisition group which received 30 training trials was included

EXPERIMENT II

Method

Three experimental groups were defined in terms of whether initial training was single-cue (S) or differential (D) conditioning, and the number of positively reinforced acquisition trials (15 or 30): D15, S15, S30. There were minor differences between the instrumentation of Exp. I and Exp. II. More importantly, Exp. II differed from Exp. I in that: (*a*) *S*s were 48 right-handed female high school students from Princeton, New Jersey. Sixteen *S*s were randomly assigned to each of the three major treatments. All *S*s were paid for participation in the experiment and were given the evaluative instructions (I) of Exp. I. (*b*) In the single-cue training conditions, half of the *S*s received 15 training trials with the Good (G) words or the Bad (B) words reinforced, Groups GS15 and BS15, respectively, and half of the *S*s received 30 training trials, Groups GS30 and BS30, respectively. The discrimination training procedures, GD15 and BD15, corresponded to Cond. GDI and BDI of Exp. I. (*c*) The stimuli used in Exp. II were high-frequency words from the semantic differential profiles of Heise (1965). The Good stimuli had average ratings of 1.70 to 2.70 on the Heise Good–Bad scale. The Neutral stimuli had average ratings of − .50 to .50, and the Bad stimuli had average ratings of − 1.70 to − 2.70. During training, the Good words were as follows: BEAUTY, FAMILY, FRIEND, NICE, PEACE, and the Bad words were: DESTROY, FAIL, HURT, SHOOT, WORRY. (*d*) The generalization test was the same as in Exp. I, except that the generalization test words were: Good—FREEDOM, LOVE, MUSIC, PLEASANT, SUCCESSFUL; Neutral—DAY, HOUR, MONTH, WEEK, YEAR; Bad—ATTACK, BURN, DEAD, FEAR, SUFFER. All the Neutral words obviously designated a period of time.

Results and Discussion

As in Exp. I, the anticipatory button-pressing response was learned equally well under all training conditions. On the last five trials of discrimination training, differential responding to the left- and right-hand buttons was highly significant, F (1, 14) = 504.79, $p < .001$. Performance on the generalization test trials is summarized in the lower panel of Fig. 1. Although the form of the generalization gradients in Exp. II differ somewhat from those obtained in Exp. I, it appears that semantic generalization along the bipolar dimension of connotative meaning was obtained. The Neutral words, belonging to a common conceptual class, produced a level of responding intermediate to that produced by the polar Good and Bad stimuli.

A most interesting difference between the response levels to the unrelated Neutral words of Exp. I and the conceptually related Neutral words of Exp. II is readily apparent in the two panels of Fig. 1. In Exp. I, the generalization gradients are very nearly linear with a slight tendency for the Neutral words to produce a response level more like that obtained with the Bad words than that obtained with the Good words. In Exp. II, however, the conceptually related Neutral words produce response levels much more like that produced by the Good words, regardless of whether the Good or Bad words had been reinforced during training. One could say that the Neutral words tended to be assimilated into the Good category. This assimilation effect is reflected in a highly significant difference in the quadratic component of the generalization trend between those *S*s for whom the Good words were reinforced and those *S*s for whom the Bad words were reinforced, F (1, 42) = 26.17, $p < .001$, $MS_e = 1.91$.

For further statistical analyses, the word-class variable was again rearranged into semantic-distance-from-the-reinforced-words variable. Overall, the semantic-distance effect was highly significant, F (2, 84) = 112.96, $p < .001$, $MS_e = 1.64$. There was a large linear component in the generalization gradient, F (1, 42) = 268.74, $p < .001$, $MS_e = 1.37$. Subsequent two-tailed t tests indicated that all three points (near, middle, and far) were significantly different from each other, showing that overall response level to the Neutral words gave an intermediate point in the gradient. The *S*s who had initially been trained to respond to Good words gave more total responses than those trained to Bad words, F (1, 42) = 19.97, $p < .001$, $MS_e = 1.38$.

The three different training conditions, 15 single-cue trials, 30 single-cue trials, and the differential reinforcement, produced no significant differences in overall responding, F (2, 42) = .52, $p > .05$, $MS_e = 1.38$. In contrast to the findings in Exp. I, the interaction of Training Condition × Semantic Distance was not significant, $F(4, 84) = 1.03, p > .05, MS_e = 1.64$. This means that although the gradient for the D15 conditions appears somewhat steeper than those for the S15 and S30 conditions, the differences were not significant, so that the significance of the differences in slope of the gradients between the single-cue and discrimination training treatments that were obtained in Exp. I were not replicated in Exp. II.

It is interesting that in spite of the great differences in the population of *S*s between the two experiments and the minor differences in procedures, the error terms for both experiments are remarkably small and very similar in magnitude, so that in a very real sense Exp. II could be called a good replication of Exp. I. In particular, the error terms for the between-*S*s effects were similar (1.71 vs. 1.38 for Exp. I and II, respectively), as were those for the overall within-*S*s effects (1.64 vs. 1.16 for the two experiments). Further, the error terms for the linear components of trend were virtually identical (1.39 vs. 1.37 for Exp. I and II, respectively), while the error terms for the quadratic components of trend differed somewhat (.93 vs. 1.91 for the two experiments).

General Discussion

In both Exp. I and Exp. II, statistically significant generalization gradients were obtained along the bipolar Good–Bad dimension of connotative meaning. These generalization gradients were obtained following both single-cue and differential training, and they possessed intermediate points which differed significantly from the response levels at the polar end points. Apparently with this type of semantic relationship among words, generalization is continuous along the Good–Bad dimension.

Although differential training produced a significantly steeper generalization gradient than did single-cue conditioning in Exp. I ($p < .05$), the actual difference in slopes was very small, and such differences as were obtained in Exp. II were not statistically significant. It appears that with this type of semantic generalization, discrimination training, as compared with single-cue training, has less effect in reducing generalization than is the case in primary stimulus generalization with nonverbal stimuli (e.g., Hoving, 1963; Reingold & Perkins, 1955).

The results of these experiments show that defining the shape or curvature of a semantic generalization gradient is a problem that is far more complex than defining the shape of the generalization gradient for nonverbal stimuli. With nonverbal stimuli, the stimulus dimension itself usually has a simple invariant physical scale, such as brightness, wavelength, frequency, sound pressure level, or the like. That semantic differential ratings of words do not provide a stimulus scale that remains invariant in different experimental contexts is evidenced by the differences in the shapes of the gradients obtained in Exp. I and Exp. II. Without stable abscissa values it is impossible to define the generalization function precisely.

The procedures used in obtaining the semantic differential ratings of words differ markedly from the way the words are employed in a conditioning experiment; because of this, it would not be surprising if the "effective" semantic differential scale placement of the words differed somewhat in the two procedures. The averaged semantic differential ratings reported by Jenkins et al. (1958) and by Heise (1965) were obtained by having each *S* rate 20 or 50 different words on 20 or 8 semantic dimensions. For each *S*, these words varied considerably on the Good–Bad dimension. In most conditioning experiments each word appears several times, and all training words are selected from the extremes of the bipolar dimension, so that during training *S* is exposed to a group of words from one pole or the other or both of the Good–Bad continuum. Thus, the interword relationship in the rating and conditioning procedures differ in such a way that different kinds of context effects should be expected.

In addition to general context effects, there was a strong apparent assimilation effect in that the response level to the conceptually related Neutral words in Exp. II was closer to the response level to the Good words than to the response level to the Bad words, regardless of which extreme of the Good–Bad continuum was reinforced. This finding in Exp. II was not expected on either an a priori basis or on the basis of the results obtained in Exp. I. The high degree of consistency in the response levels to the Neutral words in the six curves in Exp. II suggests that the assimilation effect should be easy to replicate. Because a comparable assimilation effect did not occur in Exp. I, it would appear that the conceptual relation among the Neutral words of Exp. II was a factor in producing the effect. The assimilation effect is very interesting, but the generality of the phenomenon is at present unknown. For example, will assimilation occur only with this particular set of conceptually related Neutral words or would any conceptual relationship among any group of Neutral words result in assimilation? Also, could similar assimilation effects occur in other bipolar dimensions of connotative meaning? Until these and other questions are answered there seems little point in attempting to explain the phenomenon.

Finally, it should be noted that considerable pilot work was necessary before a successful procedure was devised. Initially, a single response key and a single reinforcement light were tried. Under these conditions, however, *S*s rarely responded to the test stimuli. Progressive changes in equipment and instructions led to the present version of training and testing procedures. With two lights and two buttons, and instructions that the lights might not operate sometimes—which seemed to be very important in getting *S* to respond during the test trials—*S* can logically be required to respond on each trial. The symmetry of the response buttons and the reinforcement lights as well as the detailed instructions to *S*s appear to provide an excellent base procedure for obtaining semantic generalization.

REFERENCES

ACKER, L. E., & EDWARDS, A. E. Transfer of vasoconstriction over a bipolar meaning dimension. *Journal of Experimental Psychology*, 1964, **67**, 1–6.

BUSS, A. H. Stimulus generalization and aggressive verbal stimuli. *Journal of Experimental Psychology*, 1961, **61**, 469–473.

COLE, S., & WILLIAMS, R. L. Semantic generalization as a function of associative value of stimuli. *Psychonomic Science*, 1966, **6**, 173–174.

CRAMER, P. Magnitude and selectivity as independent factors in semantic generalization. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 509–524. (a)

CRAMER, P. Semantic generalization: IAR locus and instructions. *Journal of Experimental Psychology*, 1970, **83**, 266–273. (b)

CRAMER, P. Semantic generalization: Demonstration of an associative gradient. *Journal of Experimental Psychology*, 1970, **84**, 164–172. (c)

DICKEN, C. F. Connotative meaning as a determinant of stimulus generalization. *Psychological Monographs*, 1961, **75** (1, Whole No. 505).

FEATHER, B. W. Semantic generalization of classically conditioned responses: A review. *Psychological Bulletin*, 1965, **63**, 425–441.

GRANT, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychological Bulletin*, 1956, **53**, 141–154.

HARTMAN, T. F. Dynamic transmission, elective generalization, and semantic conditioning. In W. F. Prokasy (Ed.), *Classical conditioning: A symposium*. New York: Appleton-Century-Crofts, 1965.

HEISE, D. R. Semantic differential profiles for 1000 most frequent English words. *Psychological Monographs*, 1965, **79** (8, Whole No. 601).

HOVING, K. L. Influence of type of discrimination training on generalization. *Journal of Experimental Psychology*, 1963, **66**, 514–520.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

JENKINS, H., & HARRISON, R. Effect of discrimination training on auditory generalization. *Journal of Experimental Psychology*, 1960, **59**, 246–253.

JENKINS, J. J., RUSSELL, W. A., & SUCI, G. J. An atlas of semantic profiles for 360 words. *American Journal of Psychology*, 1958, **71**, 688–699.

KURCZ, I. Semantic and phonetographic generalization of a voluntary response. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 261–268.

LANG, P. J., GEER, J., & HNATIOW, M. The semantic generalization of conditioned autonomic responses. *Journal of Experimental Psychology*, 1963, **65**, 552–558.

LASHLEY, K. S., & WADE, M. The Pavlovian theory of generalization. *Psychological Review*, 1946, **53**, 72–87.

LEVY, C. M., & MURPHY, P. H., III. The effects of alcohol on semantic and phonetographic generalization. *Psychonomic Science*, 1966, **4**, 205–206.

MALTZMAN, I. Theoretical conceptions of semantic conditioning and generalization. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

MALTZMAN, I., & LANGDON, B. Semantic generalization of the GSR as a function of semantic distance or the orienting reflex. *Journal of Experimental Psychology*, 1969, **80**, 289–294.

MEDNICK, M. T. Mediated generalization and the incubation effect as a function of manifest anxiety. *Journal of Abnormal and Social Psychology*, 1957, **55**, 315–321.

MINK, W. D. Semantic generalization as related to word association. *Psychological Reports*, 1963, **12**, 59–67.

NOVAK, D. J., & GRANT, D. A. Semantic generalization of the conditioned eyelid response with two types of training. *Psychological Reports*, 1965, **17**, 731–738.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana, Ill.: University of Illinois Press, 1957.

PEASTREL, A. L. Studies in efficiency: Semantic generalization in schizophrenia. *Journal of Abnormal and Social Psychology*, 1964, **69**, 444–449.

PERKINS, C. C., HERSHBERGER, W. A., & WEYANT, R. G. Difficulty of a discrimination as a determiner of subsequent generalization along another dimension. *Journal of Experimental Psychology*, 1959, **57**, 181–186.

PERRY, L. C., ORNSTEIN, P. A., & GRANT, D. A. A test of associative generalization of the conditioned eyelid response. *Psychonomic Science*, 1971, **23**, 319–320.

RAZRAN, G. A quantitative study of meaning by a conditioned salivary technique (semantic conditioning). *Science*, 1939, **90**, 89–91.

RAZRAN, G. The observable unconscious and the inferable conscious in current Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychological Review*, 1961, **68**, 81–147.

REINGOLD, D. B., & PERKINS, C. C. Stimulus generalization following different methods of training. *Journal of Experimental Psychology*, 1955, **49**, 423–427.

REISS, B. T. Genetic changes in semantic conditioning. *Journal of Experimental Psychology*, 1946, **36**, 143–152.

RICE, U. M., & DIVESTA, F. J. A developmental study of semantic and phonetic generalization in paired-associate learning. *Child Development*, 1965, **36**, 721–730.

SHVARTS, L. A. Conditioned reflexes to verbal stimuli. *Voprosy Psikhologii*, 1960, **6**, 86–98. Cited by G. Razran, The observable unconscious and the inferable conscious in current Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychological Review*, 1961, **68**, 81–147.

SPENCE, K. The nature of discrimination learning in animals. *Psychological Review*, 1936, **43**, 427–449.

VOLKOVA, V. D. On certain characteristics of the formation of conditioned reflexes to speech stimuli in children. *Fiziol. Zh. SSSR*, 1953, **39**, 540–548. Cited by G. Razran, The observable unconscious and the inferable conscious in current Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychological Review*, 1961, **68**, 81–147.

VOLKOVA, V. D. On the correcting influence of the environment in elaborating conditioned reflexes to several verbal stimuli. *Zhurnal Vysshei Nervnoi Deiatel'nosti*, 1957, **7**, 525–533. (*Psychological Abstracts*, 1958 **32**, No. 5185).

(Received December 31, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 211–214

PROACTIVE EFFECTS IN SHORT-TERM RECOGNITION MEMORY[1]

DAVID S. GORFEIN[2] AND DAVID E. JACOBSON[3]

New College

The basic experimental design tested to see whether semantic shifts and variation of intertrial interval influenced performance in short-term recognition memory in a manner similar to their influence in short-term recall. The results are quite unequivocal. There is no evidence for any qualitative difference between the recognition process and the recall process in the Brown-Peterson paradigm. If the commonly held belief that proactive inhibition is due to retrieval failures is accurate, the fact that processes that influence proactive inhibition in recall also do so in recognition raise some problems for Kintsch's postulation of a two-process theory of recognition and recall.

There has been a great resurgence of interest in recognition memory in the last several years partially based upon a two-process theory of memory that suggests that recall consists of a retrieval phase plus a recognition phase whereas recognition bypasses the retrieval phase and only trace evaluation occurs (cf. Kintsch 1970). Tulving and Thomson (1971) have recently presented some evidence that indicates that retrieval is a necessary process in at least one recognition memory paradigm.

Short-term memory in the Brown-Peterson paradigm (Brown, 1958, Peterson & Peterson, 1959) has been shown to be very sensitive to proactive inhibition (PI) effects (Keppel & Underwood 1962, Wickens, Born, & Allen 1963). It is generally agreed (cf. Keppel, 1968) that PI results from a failure in retrieval. The present study therefore investigates whether the factors that influence the rate of PI buildup and dissipation in the standard short-term paradigm will have a comparable influence in short-term recognition memory. Specifically, the two factors manipulated are length of intertrial interval (ITI) and shift of semantic categories. PI has been shown to decrease as a function of the length of an ITI in the Brown-Peterson paradigm (Loess & Waugh, 1967; Peterson & Gentile, 1965). In the Loess & Waugh study, when the ITI exceeded 2 min. proactive inhibition was found to be negligible. By shifting from consonant triads to number triads, Wickens et al. were able to obtain release on the first item of a new set. In a similar manner Loess (1968) was able to show repeated release from proactive inhibition by changing semantic categories.

The present study therefore tests to see if short-term recognition memory in the Brown-Peterson paradigm is influenced by the same factors that effect PI buildup in recall in short-term memory. Should this be the case, and if the commonly held belief that proactive inhibition is due to retrieval failures is correct, we would have demonstrated another failure for Kintsch's two-process theory of memory.

EXPERIMENT I

Method

Subjects.—The *S*s were 36 New College students who volunteered to serve in the experiment.

Procedure.—Items were presented with a Kodak Carousel projector employing a 1.5-sec. ready signal, a 1.1-sec. item presentation, a 10-sec. retention period, a 10-sec. recognition period, and varying rest periods (intertrial intervals). The retention intervals were filled by a Stroop color-naming task. Items were presented in blocks of six trials with consonant and digit trigrams in alternating blocks for a total of 12 blocks. Within a block, ITIs were either 5, 30, or 90 sec., with 30 sec. separating blocks. The serial position for items within blocks was counterbalanced using a 6 × 6 Latin square. The order of ITIs within the experiment was similarly counterbalanced with a Latin square. At the onset of the recognition period, *S*s turned over a 3 × 5 in.

---

[1] This is an extended version of a paper by the same title presented at the Midwestern Psychological Association meetings in Detroit, 1971.

[2] Requests for reprints should be sent to David S. Gorfein, Division of Natural Sciences, New College, Sarasota, Florida 33578.

[3] This paper represents part of a senior thesis conducted by the second author under the direction of the first author.

white card and chose 1 of 6 alternatives typed on the card. Alternatives for CCCs included the correct item, two acoustic confusions based on it (cf. Pearthree, Gorfein, & Bennett, 1967), the previous item, and two acoustic confusions based on the previous item. Alternatives for the digits included the correct item, two permutations of it, the previous item, and two permutations of it. In the case of the first item in a block, a dummy previous item was used. Testing was individual with *S* seated in a sound-attenuating cubicle. His performance, including the Stroop color-naming task was monitored by *E* over an intercom system.

Materials.—The six blocks of consonants consisted of 50% association value trigrams (Witmer, 1935). The digit trigram consisted of three digit numbers with the restriction that the first digit was greater than zero and no digit was repeated within a trigram. Items were so arranged that within a block of six trials no consonant or digit occurred more than once in the same within-trigram serial position.

Results and Discussion

The experimental design allowed for only one observation at each serial position within each of the 12 alternating categories of materials. Therefore, in order to test whether the effect of semantic shifts in the group produced repeated release from PI, we looked at the first position in each category as compared to the average of the other five positions across the 12 blocks. In all 12 blocks the first position was higher than the average of the other positions, thus showing that the semantic shifts did produce repeated PI release.

Figure 1 shows the serial position effects as a function of intertrial interval (ITI) collapsed over the two types of material. While the data tends to be noisy, statistical analysis shows that the effect of ITI is similar to that obtained in the recall situation by Loess and Waugh (1967). With the data collapsed across materials, the ITI × Serial Position interaction is significant, $F\ (10, 350) = 2.19$, $p < .05$. The main effects of ITI and serial position are both highly significant; for ITI, $F\ (2, 350) = 6.99$, $p < .01$, and for serial position, $F\ (5, 350) = 12.14$, $p < .001$. We would expect the significant interaction effect, based on the PI buildup function as the first position in the category should be substantially the same for all ITIs and then PI should buildup at different rates.

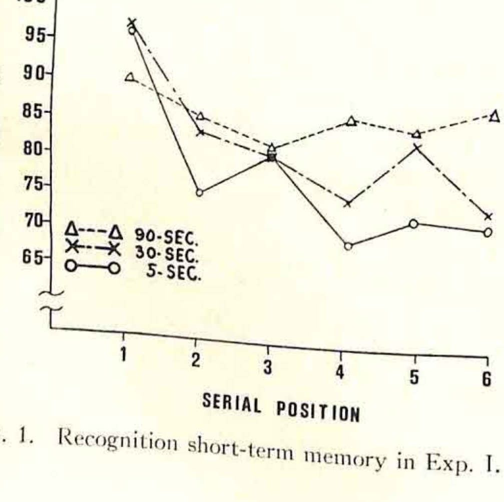


FIG. 1. Recognition short-term memory in Exp. I.

An analysis was made of the type of errors that *S*s made in the recognition task. We analyzed to see whether when *S* made an error he picked a permutation of the current item or chose the previous item. We will reserve discussion of this data until the report of Exp. III, which contains some control information.

EXPERIMENT II

While we were able to obtain the expected effects of ITI in the first study, the noise in the data and the failure to get complete PI release due to ITI led us to repeat the study using a longer intertrial interval.

Method

Subjects.—Twelve New College students served as *S*s in this second study.

Procedure.—The procedure and materials for the second study were identical to that of the first except that we used only two ITIs, a 5-sec. rest period, and a 150-sec. rest period. Items and ITIs were again suitably counterbalanced.

Results and Discussion

Figure 2 shows the serial position function for the two sets of ITIs. It is clear from the figure that the 150-sec. ITI results in negligible, if any, PI buildup while the 5-sec. ITI shows the same effects as in our first experiment. Statistical analysis shows that for these data, i.e., with the data collapsed across materials, there is a significant ITI × Serial Position interaction effect with $F\ (5, 55) = 3.17$, $p < .05$. The main

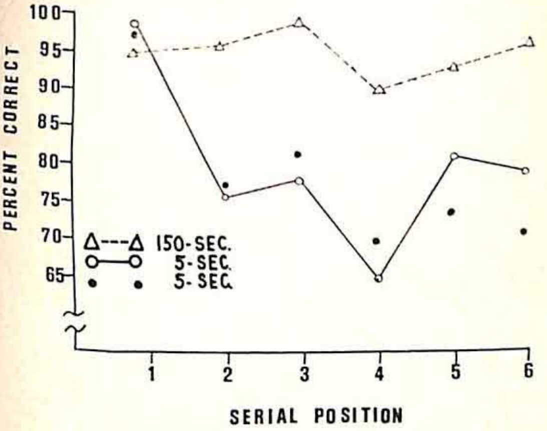


FIG. 2. Recognition short-term memory in Exp. II.

effect of serial position was significant, $F(5, 55) = 5.53$, $p < .001$. The effect of intertrial interval also showed a large and significant difference, $F(1, 55) = 26.69$, $p < .001$. In order to ascertain whether the 150-sec. ITI had indeed produced a lack of PI buildup we analyzed the data separately for the 150-sec. condition, using a two-way Ss serial position analysis. This analysis confirmed that there was no significant serial position effect in the 150-sec. ITI condition, $F(5, 55) = .79$.

EXPERIMENT III

In Exp. I and II, three of the six alternatives on the multiple-choice card were the previous item and two variations of it. On the first member of a semantic set, a dummy item was used in lieu of the last item of the previous semantic class (This could, of course, have influenced the guessing probabilities as it is possible that the member of the other semantic class, the previous item, might have been a better alternative, although this is highly doubtful.) A much more crucial question about the role of the three alternatives from the prior item is whether they have their effect as distractors because they are the prior item or because *S* does not recognize the current item and then chooses at random among the alternatives. The third study therefore is designed to test the effectiveness of distractors that come from prior items as compared to the use of irrelevant distractors.

TABLE 1

ERROR TYPES IN EXPERIMENT I AS FUNCTION OF WITHIN-CATEGORY SERIAL POSITION

Position	Variation of current item	Previous item	Omissions	Total
1	22	4	0	26
2	32	43	2	77
3	46	33	5	84
4	53	46	1	100
5	50	40	3	93
6	61	30	6	97
Total	264	196	17	477

Method

Subjects.—Twelve New College students served as *Ss* in this third study.

Materials and procedure.[4]—The items employed were the same 36 CCCs and 36 NNNs employed in Study 1. For half of the *Ss*, half the items of each material chosen at random were given recognition alternatives that contained a current item and a previous item as in Exp. I. The remaining half of the items had recognition alternatives which contained the current item and its related distractors as in Exp. I and a dummy item (one with no letter or digit in a serial position common with the to-be-remembered item, or the prior item) and two alternative forms of the dummy. For the other half of the *Ss*, the items were shifted so that items which previously had dummy alternatives now had prior item alternatives and vice versa. Items were presented with only a 5-sec. ITI in order to maximize the number of errors we would get for analysis. All other aspects of the procedure were the same as in Exp. I.

Results and Discussion

As we mentioned above, the main purpose of Exp. III was to give us a base line of comparison for Exp. I. Table 1 shows an analysis of the types of errors in Exp. I as a function of the serial position within the six-item block (the error analysis includes all ITIs and both types of material). In Serial Positions 2–6 (where true previous item intrusions are possible), there are a total of 242 current item errors as compared to 192 prior item errors. Since there are only two incorrect alternatives based on the current item and three incorrect alternatives based on the prior item, the data

---

[4] The authors are indebted to Lynne Tarakan who prepared the dummy items needed for this study and to Craig Evinger who served as *E* in Exp. III.

TABLE 2

ANALYSIS OF ERRORS AS A FUNCTION OF ALTERNATIVES PRESENT (EXPERIMENT III)

Cond.	Error type for Positions 2-6			
	Variation of current item	Intrusion	Omissions	Total
Previous present	38	24	2	64
Dummy present	48	13	1	62

clearly suggest that all errors are not pure guesses. It is conceivable that *S* is making use of partial information when he chooses among the alternatives. This will suggest the hypothesis that part of the difference in overall performance between recognition and recall in most studies is due to *S*s ability to make use of partial information in recognition.

Table 2 shows the analysis of errors as a function of the type of alternative present in Exp. III. The crucial data to note in this table is that *S*s make approximately the same number of errors when the alternatives contain a dummy item as they do when they contain the prior item. Incorrect variations of the current item are chosen in either case more frequently than any of the three variations of the prior or dummy item. There seems to be a strong implication from these data that whatever prior items do to produce PI, having them present as recognition alternatives has no specific influence on the likelihood of an error.

The combined results of all three studies show unequivocally that recognition in the multiple-choice test employed produces results that do not differ qualitatively from recall, in terms of the influence of PI in the Brown-Peterson paradigm. No support is found in this data for the Kintsch (1970) two-process theory of memory. Instead, added support is given to the position that recognition differs quantitatively from recall due to the fact that *S* can make greater use of partial information in the recognition situation.

REFERENCES

BROWN, J. A. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 1958, **10**, 12–21.

KEPPEL, G. Retroactive and proactive inhibition. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs: Prentice-Hall, 1968.

KEPPEL, G., & UNDERWOOD, B. J. Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 153–161.

KINTSCH, W. Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

LOESS, H. Short-term memory and item similarity. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 87–91.

LOESS, H., & WAUGH, N. C. Short-term memory and intertrial interval. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 455–460.

PETERSON, L. R., & GENTILE, A. Proactive interference as a function of time between tests. *Journal of Experimental Psychology*, 1965, **70**, 473–478.

PETERSON, L. P., & PETERSON, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

PEARTHREE, V., GORFEIN, D. S., & BENNETT, R. W. Acoustic data and recognition in short-term memory. *Psychonomic Science*, 1967, **8**, 59–60.

TULVING, E., & THOMSON, D. M. Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, 1971, **87**, 116–124.

WICKENS, D. D., BORN, D. G., & ALLEN, C. K. Proactive inhibition and item similarity in short-term memory, *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 440–445.

WITMER, L. R. The association value of three-place consonant syllables. *Journal of Genetic Psychology*, 1935, **47**, 337–360.

(Received December 31, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 215–222

MAGNITUDE JUDGMENTS OF BRIGHTNESS AND BRIGHTNESS DIFFERENCE AS A FUNCTION OF BACKGROUND REFLECTANCE[1]

DWIGHT W. CURTIS[2]
California State College at Fullerton

STANLEY J. RULE
University of Alberta

Magnitude estimates of the brightness of Munsell grays and estimates of difference between paired grays were obtained from 48 *Ss*, employing backgrounds with high and low reflectances, and using two different standard stimuli. Difference estimates were analyzed to yield scales of apparent brightness and to determine the form of an output transformation employed by *Ss* in assigning numbers to stimuli. Both the output functions and the functions relating scale values to reflectance (input functions) were well-approximated by power functions. The exponent of the input function from the high-reflectance background condition was approximately twice as large as the exponent of the same function from the low-reflectance background condition. The product of the input and output exponents was found to predict magnitude estimates of brightness reasonably well in the high-reflectance background condition, but not in the low-reflectance background condition. Prediction was improved in the latter case, however, by the introduction of an additive constant, translated on the stimulus axis, whose value was the reflectance of the background.

S. S. Stevens and his associates have used magnitude estimation procedures to investigate the psychophysical relation between perceived brightness and stimulus luminance and found the relation to be well-approximated by a power function,

$$J = a\phi^n. \qquad [1]$$

For the dark-adapted eye, the relation was approximately a cube-root function (Stevens & Galanter, 1957). As the eye adapted to higher levels of illumination, however, the size of the exponent was found to increase (Stevens & Stevens, 1963). Presentation of stimuli in the context of a brighter inducing field was also found to increase the magnitude of the exponent (Horeman, 1965; Stevens & Diamond, 1965). For lightness judgments of gray papers, a higher exponent of 1.2 was found (Stevens & Galanter, 1957).

The form of the power function in Equation 1 does not provide as satisfactory an approximation of the empirical relation in the light-adapted state, owing to systematic deviations as the function approaches the origin. Stevens and Stevens (1963) corrected for these deviations by employing a more general form of the function,

$$J = a(\phi - c)^n, \qquad [2]$$

where the constant c was interpreted as the "effective threshold." Both S. S. Steven's use of the constant and his interpretation of it have been criticized (e.g., Corso, 1963). Fagot and Stewart (1968) have favored an alternative function,

$$J = a(\phi^n - \phi_0^n),$$

which introduces a threshold constant, ϕ_0, by translation on the response axis.[3] A function,

$$J = a\phi^n + b, \qquad [3]$$

used by Curtis (1970) is formally equivalent, although the latter investigator interpreted the additive constant b as a response parameter.

Magnitude judgment instructions may be used to scale differences, as well as single stimuli. Curtis, Attneave, and Harrington (1968) and Curtis (1970) obtained evidence that judgments of apparent difference and apparent magnitude constitute independent operations which, by their convergence, support a two-stage model of magnitude judgment. The model was based on an

---

[1] This research was supported by Grant APA-151 from the National Research Council of Canada.

[2] Requests for reprints should be sent to Dwight W. Curtis, Department of Psychology, California State College, Fullerton, California 92631.

[3] See Marks and Stevens (1968) for a more detailed discussion of the controversy concerning the additive constant.

earlier suggestion by Attneave (1962) that magnitude estimation confounded the sensory function with an output function having to do with the psychological significance of number. It was assumed that each stimulus intensity induces in O a psychological correlate, ψ, related to physical intensity by a power function,

$$\psi = a\phi^k,$$

and that the psychological correlate of a number is also related to objective number by a power function. Consequently, the task of S in a magnitude estimation experiment is to match the psychological correlate of a stimulus with the psychological correlate of a number. If such is the case, the empirical power function (Equation 1) relating numerical responses to stimulus intensity may be expressed as

$$J = a\phi^{km}, \qquad [4]$$

where k is an input or sensory parameter and m is an output parameter, the reciprocal of the exponent for subjective number. From Attneave's model, Curtis et al. (1968) derived the expression

$$J_{ij} = a(\phi_i{}^k - \phi_j{}^k)^m + b \qquad [5]$$

for judgments of differences. Equation 5 was found to provide an excellent fit to the data. However, it provided no basis for independently examining the forms of the input and output transformations: the relation between psychological magnitude and stimulus magnitude and the relation between judgments and psychological differences. Such an examination was provided by Rule, Curtis, and Markley (1970) using a nonmetric scaling approach. They assumed that each stimulus magnitude, ϕ, is associated with a central process, with a scale value, V, such that $V = f(\phi)$ denotes the input transformation. If Ss difference judgments are a function of differences between the central processes associated with two stimuli, numerical assignments to differences may be expressed as

$$J_{ij} = g(V_i - V_j).$$

Rule et al. (1970) solved for V under the assumption the output transformation, denoted by g, was monotone. The output transformation determined from the relation between J_{ij} and $(V_i - V_j)$ was found to be a power function, which may be expressed as

$$J_{ij} = a'(V_i - V_j)^m + b'. \qquad [6]$$

The input transformations for the stimuli they investigated, circle area and weight, were found to be power functions of the form

$$V = c\phi^k + d. \qquad [7]$$

The additive constant, d, was necessary since the scale values were on an interval scale, and hence the origin was arbitrary.

In the present study, a similar approach was used to investigate the form of the input transformation for brightness under conditions of adaptation and contrast. However, a direct solution of Equation 6, rather than the nonmetric technique, was used to directly estimate the scale values, V, the exponent m, and the additive constant, b, and the form of the input relation was recovered from the relation between scale values and physical reflectance.

METHOD

Seven stimuli were used in the experiment, consisting of 7.62 × 12.7 cm. rectangular papers from the Munsell neutral value scale, mounted on thin cardboard. The Munsell values of the stimuli ranged from 2.5 to 8.5 in unit intervals. Their reflectances were .046, .090, .156, .246, .362, .507, and .684.

The stimuli were presented to S on a 76.2 × 116.8 cm. table, the surface of which was covered by heavy cardboard constituting the background for the stimuli. Two background conditions were used in the experiment. In the low-reflectance background condition the reflectance of the background was .028, and in the high-reflectance background condition the background reflectance was .946.

The Ss were 48 students who participated in the experiment to fulfill a requirement of an introductory psychology course at the University of Alberta. Twelve Ss were assighed at random to each of four groups. Each S performed two tasks, magnitude estimation of brightness (lightness) and magnitude estimation of difference. Two of the groups performed the tasks with the high-reflectance background and the other two groups performed the same tasks with the low-reflectance background. The two groups within each background condition received different standard stimuli and standard differences. For one group, the stimulus with the highest Munsell value served as standard in estimating brightness, and the largest difference between Munsell values served as standard in estimating differences. For the other group, the standards were the stimulus with the lowest Munsell value for estimating brightness and the difference between the two stimuli with the lowest Munsell values for estimating differences.

RESULTS AND DISCUSSION

A direct solution of Equation 6 was not possible since the coefficient a and scale values, V, were not independent. Consequently, the equation

$$J_{ij} = (V_j' - V_i')^m + b \qquad [8]$$

was fitted to the difference estimates. In Equation 8, each $V_i' = a^{1/m} V_i$. Estimates of V', the output exponent m, and the additive constant b were attained by a least-squares solution of

$$J_{ij} = (X_1 V_1' + X_2 V_2' + \cdots + X_r V_r')^m + b,$$

where the X_j were assigned values 1, 0, or -1. For instance, for the judged difference between Stimulus 2 and Stimulus 5, $X_2 = -1$, $X_5 = 1$, and all other X_j were zero. A solution for Equation 8 was achieved by an iterative procedure minimizing Q, where

$$Q = \sum_{i=1}^{n} W_i [Y_i - f(X_1, X_2, \ldots, X_r, V_1, V_2, \ldots, V_r, m, b)]^2 \qquad [9]$$

In Equation 9, Y_i is the dependent measure, the geometric mean judgment of the ith difference, the X_j are independent variables, and the V_j, m and b are parameters to be estimated. The symbol W_i refers to a weight assigned to the ith observation. By letting $W_i = 1/Y_i^2$ one minimizes the sum of squares of the relative deviations, a strategy approximately comparable to a least-squares analysis performed on the logarithms of Y. When $W = 1$, all deviations are weighted equally. The first weighting strategy is the more consistent with customary practice and with the usual finding that variability increases as a function of stimulus (and response) magnitude. Curtis and Fox (1969) found, however, that homoscedasticity was more closely approximated by the distribution of magnitude judgments than by the distribution of their logarithms when the stimulus used as standard was greater in magnitude than the stimuli with which it was compared. The present analyses employed weights of 1 in fitting functions to data from those conditions with a high standard and the weighting $1/Y_i^2$ in those conditions employing a low-standard stimulus.

High-reflectance background.—For the two groups that received the high-reflectance background a least-squares fit of Equation 8 to the geometric means of difference judgments yielded

$$J_{ij} = (V_i - V_j)^{1.49} + .61$$

for the High-Standard condition, and

$$J_{ij} = (V_i - V_j)^{1.20} + 3.65,$$

for the Low-Standard condition. For each group a plot of judged difference against difference between scale values raised to the mth power is presented in Fig. 1. Owing to differences in the weights applied in the least-squares solutions, no significance should be attached to the difference in output exponents. When the same weights were applied in both solutions, the difference was substantially reduced.

Equation 7 was fitted to the relation between scale values, V, and physical reflectance. A least-squares solution, employing weights equal to unity, resulted in

$$V = 1.60\phi^{.34} - 2.63$$

for the High-Standard condition, and

$$V = 6.57\phi^{.39} - 11.64,$$

for the Low-Standard condition. The two functions are shown in Fig. 2. Both scales are approximately linear with Munsell value.

Substituting the right side of Equation 7 for scale values, V, in Equation 6 yields

$$J_{ij} = a'(\phi_j^k - \phi_i^k)^m + b, \qquad [10]$$

where $a' = c^m$. Substituting the parameter estimates obtained from the data of the High-Standard condition resulted in a stimulus–response relation of the form

$$J_{ij} = 2.01(\phi_j^{.34} - \phi_i^{.34})^{1.49} + .61.$$

The comparable function for the Low-Standard condition was

$$J_{ij} = 9.61(\phi_j^{.39} - \phi_i^{.39})^{1.20} + 3.65.$$

The difference estimates of individual *S*s were subjected to a similar analysis. Means of the input exponent, k, were .35 for the High-Standard condition, and .37 for the Low-Standard condition. The standard deviations were .14 and .18, respectively. Means of m for the two conditions were 1.58 and 1.27, with standard deviations of .63 and .46, respectively. In all cases, mean exponents are approximately of the same order of magnitude as the exponents calculated from the group data, although means of m tend to be slightly higher than their group counterparts.

Equation 10 is formally identical to Equation 5, employed by Curtis et al. (1968) and Curtis (1970) to test the two-stage model. In these previous studies, Equation 5 was fitted directly to difference-estimation data employing a least-squares solution. Comparable solutions for the present data, employing the same weights employed in the scaling solutions, yielded the functions

$$J_{ij} = 1.43(\phi_j^{.35} - \phi_i^{.35})^{1.74} + .90$$

for the High-Standard condition, and

$$J_{ij} = 9.73(\phi_j^{.38} - \phi_i^{.38})^{1.26} + 4.26$$

for the Low-Standard condition. These functions differ from those obtained from the scaling solutions largely with respect to their values of the output exponent m. The estimate of this parameter was larger from the solution for Equation 5, especially for the High-Standard data, than from the solution for Equation 6, in which the input transformation was not required to be a power function. This appears to indicate that the power function may not be a completely satisfactory approximation to the relation between scale values and stimulus intensities. The functions in Fig. 2, it may be noted, are slightly S shaped rather than linear, especially in the High-Standard condition where the discrepancy is largest.

An assumption of the two-stage model, implied in Equation 4, is that the exponent n of the power function relating magnitude estimates of single stimuli to stimulus reflectance is equal to the product of k and m, the input and output exponents, respectively. Equation 3 was fitted to judgments of brightness, resulting in a function for the High-Standard condition of

$$J = 1.09\phi^{.55} - 1.36,$$

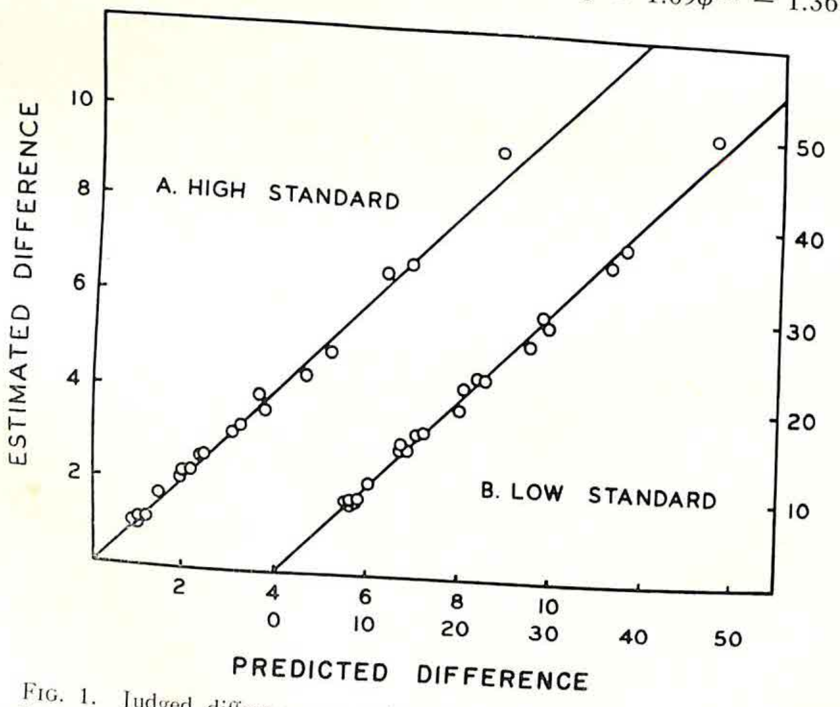


FIG. 1. Judged difference as a function of predicted difference: high-reflectance background. (Value on the abscissa are predicted by $J_{ij} = (V_j - V_i)^m + b$.)

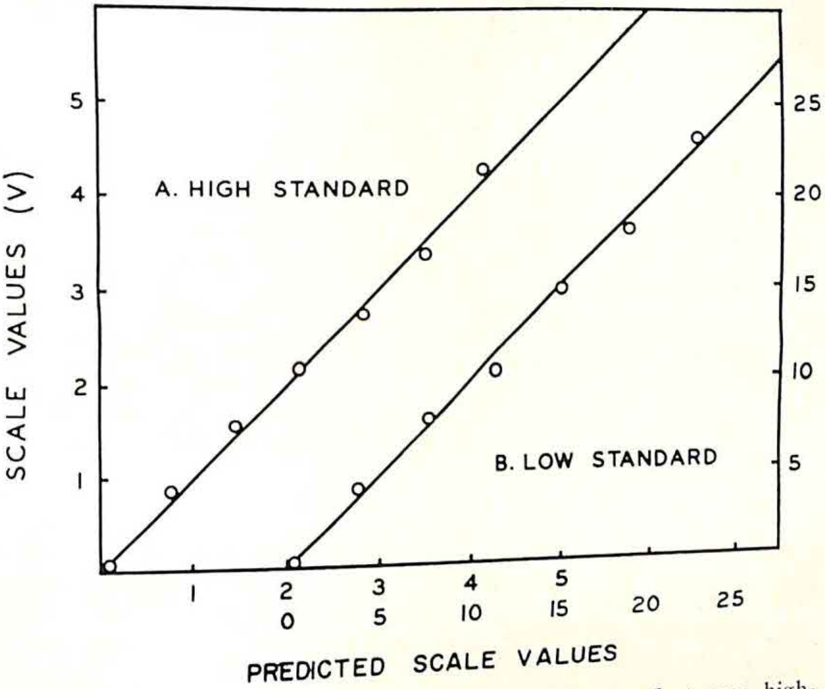


FIG. 2. Scale value (V) as a function of stimulus reflectance: high-reflectance background. (Values on the abscissa represent prediction from $V = c\phi^k + d$.)

and a function for the Low-Standard condition of

$$J = 10.98\phi^{.41} - 9.18.$$

The predicted exponents from the relation $n = k \times m$ were .51 and .47 for the High-Standard and Low-Standard conditions, respectively. The mean of the two predicted exponents was .49, which was close to the mean exponent of .48 for the magnitude estimation of single stimuli.

Mean exponents from individual *S*s' magnitude estimates were .51 for the High-Standard condition and .50 for the Low-Standard condition. The standard deviations were .48 and .44, respectively. The two means of n were more similar in magnitude than were the exponents estimated from the group data. Furthermore, the means are closer to estimates of n from the products of mean k and mean m than was true where the exponents were estimated from the pooled data: $km = .55$ for the High-Standard condition, and .48 for the Low-Standard condition.

Curtis (1970) and Curtis et al. (1968) found that the convergence of n and km held for the data of individual *S*s, as well as the group data. There was no correlation between the two sets of estimates in the present experiment. This probably indicates that the present results from individual *S*s are less reliable than were those from the previous investigations. The *S* variation was greater on all three of the exponents than was previously found, particularly in the case of n estimated from single-stimulus judgments.[4] The standard deviations of the latter exponent were twice as great as obtained by Curtis (1970). This may indicate that the brightness of grays is more difficult for *S*s to estimate than some other attributes, but it also is probable that a greater number than six observations of each of seven stimuli is required to obtain a reliable evaluation of the performance characteristics of individual *S*s.

Low-reflectance background.—For the data from the condition with a low-reflectance background, a least-squares fit of Equation

---

[4] The averaged judgments of a few *S*s in both this condition and the low-reflectance background condition exhibited more extreme curvature than either a power function with a positive exponent or a logarithmic function. For purposes of averaging, negative exponents and coefficients were permitted for these *S*s. No theoretical significance was attributed to the negative values of these exponents.

6 to the geometric means of *S*s' magnitude estimations yielded

$$J_{ij} = (V_j - V_i)^{1.46} + .39$$

and

$$J_{ij} = (V_j - V_i)^{1.50} + 2.43$$

for the High- and Low-Standard conditions, respectively. The relation between difference estimates and values predicted by these functions are shown in Fig. 3. Power functions (Equation 7) fitted by the method of least squares to the relation between scale values and reflectance were

$$V = 7.46\phi^{.15} - 9.27$$

for the High-Standard condition and

$$V = 21.21\phi^{.14} - 26.24$$

for the Low-Standard condition. These relations are shown in Fig. 4. These input exponents were very close to the exponent obtained by Curtis (1970) from dark-adapted *S*s estimating differences between luminous stimuli in total darkness.

Combining the results of the above analyses in terms of Equation 10 yielded the functions

$$J_{ij} = 18.80(\phi_j^{.15} - \phi_i^{.15})^{1.46} + .39$$

and

$$J_{ij} = 97.68(\phi_j^{.14} - \phi_i^{.14})^{1.50} + 2.43,$$

describing the relations between difference estimates and stimulus reflectance for the High- and Low-Standard conditions, respectively. Least-squares solutions of Equation 5 fitted directly to this relation yielded the functions

$$J_{ij} = 22.52(\phi_j^{.14} - \phi_i^{.14})^{1.62} + .62$$

and

$$J_{ij} = 68.62(\phi_j^{.17} - \phi_i^{.17})^{1.54} + 2.57$$

for the respective conditions.

Power functions (Equation 3) fitted to the relation between magnitude estimates of single stimuli and physical reflectance yielded exponents, *n*, of .52 and .45 for the High- and Low-Standard conditions, respectively. The mean exponent of .48 was very similar to that of .47 from the high-reflectance background condition. This result appears inconsistent with the results of Stevens and Diamond (1965), who studied the effects of the surrounding field on magnitude estimates of brightness. Furthermore, the exponents are not

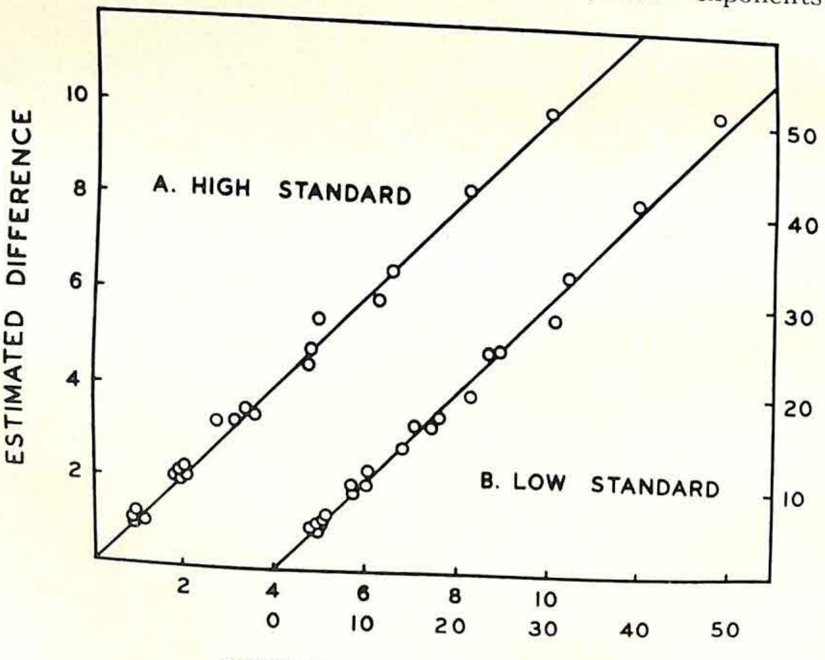


FIG. 3. Judged difference as a function of predicted difference: low-reflectance background. (Values on the abscissa are predicted by $J_{ij} = (V_j - V_i)^m + b$.)

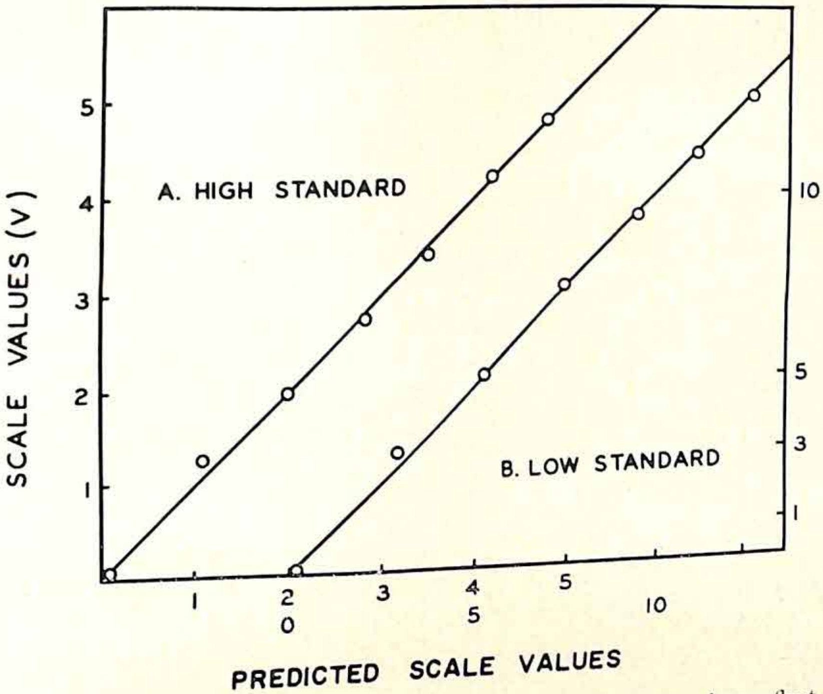


FIG. 4. Scale value as a function of stimulus reflectance: low-reflectance background. (Values on the abscissa represent predictions from $V = c\phi^k + d$.)

predicted by the hypothesized relation, $n = km$. The products for the two low-reflectance conditions—$.15 \times 1.46 = .22$, and $.14 \times 1.50 = .21$—are less than half as large as the exponents estimated from single brightness estimates.

Results of analyses of the data of individual *S*s are approximately comparable to those from the group data, although the means of k, m, and n are all slightly higher than their estimate from the pooled data. Mean values of k from the High- and Low-Standard conditions were .18 and .15, with standard deviations of .29 and .17. Means of m were 1.46 and 1.57 for the same conditions, with standard deviations of .46 and .47. Mean values of n were .56 and .55, with standard deviations of .50 and .48.

Clearly, a two-stage model in the present form does not reconcile brightness judgments with difference judgments for the low-reflectance background, although the model was consistent with data from the high-reflectance background condition. The inconsistency may have been due to different effects of background operating in the two background conditions. The high-reflectance background introduces a contrast condition, and its effect may be largely inhibitory as Stevens and Diamond (1965) suggested, influencing only the exponent of the power function. The low-reflectance background may, instead, influence *S*s' adaptation level, implicating an additive constant on the stimulus axis, as Stevens and Stevens (1963) have suggested. Hence an equation of the form

$$V = c(\phi - .028)^k + d \qquad [11]$$

was fitted to each of the scales from the low-reflectance background condition by the method of least squares, estimating c, k, and d. A constant .028, the reflectance of the background, was subtracted from each stimulus value. The values of k obtained from this analysis were .31 from the High-Standard condition and .29 from the Low-Standard condition. The effect of the additive constant was to enhance the value of k so that it approached in magnitude the exponents from the high-reflectance background conditions. The multiplicative relation, $n = km$, employing these estimates of k together with the values of m estimated from difference judgments predicted values of n of .47 and .44 for the high- and low-reflectance backgrounds, respectively. Comparable estimates of n were obtained by fitting the equation

$$J = a(\phi - .028)^n + b \qquad [12]$$

to magnitude estimates of single stimuli. Equation 12 is a modification of Equation 3, which included the background reflectance as an additive constant translated on the stimulus axis. But Equation 12 incorporates an additive constant on the response axis, as well as a threshold or adaptation constant. The latter constant was required so that the analyses of magnitude judgments of single stimuli in this background condition would be comparable to the analyses of difference judgments from the same condition and to analyses of magnitude and difference judgments from the high-reflectance background condition. Inclusion of the constant follows from the assumption that *S*s' judgments reflect a displacement of the subjective origin from zero on the response scale. Estimates of n in terms of Equation 12 were .55 and .46 for the High- and Low-Standard conditions, respectively. The slopes of the functions were .81 and 5.50, and the additive constants, b, were .29 and 3.94.

It would appear that an additive constant translated on the stimulus axis corrects the two-stage model so that convergence between the estimates of n from brightness estimation and difference estimation is more nearly approximated. According to Stevens and Stevens (1963), such a constant defines the origin of the effective stimulus scale, an origin which shifts as *S*s' level of light adaptation changes. These investigators found that the magnitudes of both the additive constant and the exponent increased as *S*s adapted to higher luminance levels. Stevens and Diamond (1965) found no need for an additive constant in the contrast experiment, although exponents of the power function were larger in this condition, and increased as the luminance of the contrasting field increased. The present results support the use of the constant in the adaptation condition; however, the magnitude of the exponent becomes very nearly equal to the exponent from the contrast condition once the correction has been applied. Fagot and Stewart (1968) have also argued for the incorporation of a threshold constant in the power function. The function they have defended, however, introduces the threshold constant by translation on the response axis. The latter constant could not have resolved the discrepancies between values of n estimated from judgments of differences and single stimuli in the present experiment.

REFERENCES

ATTNEAVE, F. Perception and related areas. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 4. New York: McGraw-Hill, 1962.

CORSO, J. F. A theoretico-historical review of the threshold concept. *Psychological Bulletin*, 1963, **60**, 356–370.

CURTIS, D. W. Magnitude estimations and category judgments of brightness and brightness intervals: A two-stage interpretation. *Journal of Experimental Psychology*, 1970, **83**, 201–208.

CURTIS, D. W., ATTNEAVE, F., & HARRINGTON, T. L. A test of a two-stage model for magnitude estimation. *Perception & Psychophysics*, 1968, **3**, 25–31.

CURTIS, D. W., & FOX, B. E. Direct quantitative judgments of sums and a two-stage model for psychophysical judgments. *Perception & Psychophysics*, 1969, **5**, 89–93.

FAGOT, R. F., & STEWART, M. R. An experimental comparison of stimulus and response translated power functions for brightness. *Perception & Psychophysics*, 1968, **3**, 297–305.

HOREMAN, H. W. Relations between brightness and luminance under induction. *Vision Research*, 1965, **5**, 331–340.

MARKS, L. E., & STEVENS, J. C. The form of the psychophysical function near threshold. *Perception & Psychophysics*, 1968, **4**, 315–318.

RULE, S. J., CURTIS, D. W., & MARKLEY, R. Input and output transformations from magnitude estimation. *Journal of Experimental Psychology*, 1970, **86**, 343–349.

STEVENS, J. C., & STEVENS, S. S. Brightness function: Effects of adaptation. *Journal of the Optical Society of America*, 1963, **53**, 375–385.

STEVENS, S. S., & DIAMOND, A. L. Effect of glare angle on the brightness function for a small target. *Vision Research*, 1965, **5**, 649–659.

STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 1957, **54**, 377–411.

(Received January 10, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 223–225

SHORT REPORTS

EFFECTS OF SOLUTION FAMILIARITY AND NUMBER OF ALTERNATIVES ON PROBLEM DIFFICULTY[1]

ROGER L. DOMINOWSKI[2]

University of Illinois at Chicago Circle

Each of 150 *S*s solved a series of 10 verbal problems requiring the selection of a "correct" word from a list of category members. Both the number of words in a list (5, 15, or 25) and the normative frequency of the correct item (five levels) were varied according to a factorial design. Errors increased as a linear function of the number of alternatives and generally increased as solution familiarity decreased, although a curvilinear effect of familiarity was also obtained. The effects of increasing the number of alternatives varied with solution familiarity in a systematic fashion. The results were interpreted in terms of an initial tendency for *S*s to guess in order of frequency followed by a change to guessing words at the appropriate frequency level.

Previous research has established that the difficulty of simple problems increases as a function of the number of alternatives (N) from which the correct alternative must be selected. The data suggest that difficulty is linearly related to N when each alternative must be separately considered (e.g., Davis, 1967), whereas difficulty is linearly related to $\log_2 N$ when the task allows the use of more efficient information-gathering strategies (Neimark & Wagner, 1964).

In these studies, the alternatives were presumably equally likely to be selected. In contrast, if the set of alternatives is hierarchically arranged, it might be expected that the effect of increasing N will vary with the position of the correct alternative in the hierarchy. More specifically, for a set of items ordered in terms of frequency (familiarity), increasing N should prove more detrimental for solutions of lower relative frequency. There is reliable evidence indicating that items of higher frequency tend to be emitted earlier than those of lower frequency (e.g., Dominowski, 1967; Duncan, 1966). Consequently, for a high-frequency solution, most of the alternatives added as N increases would be of lower frequencies and thus should not interfere with discovery of the solution; this would be less true for solutions of lower relative frequencies. Alternatively, the inverse relationship between difficulty and solution familiarity should be more pronounced as N increases.

Studies of solution familiarity with selection tasks indicate additional complexities to be considered. When *S* is given a set of items varying in frequency and is told the relative frequency of the correct item, selection difficulty is curvilinearly related to solution frequency (Duncan, 1967). In the absence of such explicit instructions, *S* will acquire this information over a series of similar problems and can thus improve his performance, especially when the solution is of low frequency (Dominowski, 1965). It should be noted that, in these studies, low-frequency solutions were quite discriminable from the alternatives in terms of frequency, in which case knowledge of the frequency of the solution can be used most efficiently (Duncan, 1967). However, typical frequency hierarchies (e.g., words belonging to a category) contain a majority of low-frequency items. Therefore, as the number of items presented increases, the discriminability of a low-frequency solution decreases, in which case it might prove more difficult to improve performance on the basis of the knowledge that the solution is a low-frequency item.

In the present study, each *S* was given a series of problems each of which required him to select the correct alternative from a set of words forming a frequency hierarchy. Both the relative frequency of the solution and N varied among groups. The major purpose of the study was to investigate the expected interaction between solution familiarity and N. In addition, changes in the effects of these variables with practice were examined.

Method.—From the Cohen, Bousfield, and Whitmarsh (1957) norms, the 25 most frequently given responses were chosen for each of 10 categories. For each category, problem sets of three sizes were constructed: For $N = 5$, the alternatives were the words occupying frequency ranks 1, 3, 6, 10, and 15 in the norms; for $N = 15$, the alternatives were the 15 most frequent words; for $N = 25$, all words chosen from the category were used.

A factorial design was used, consisting of the combinations of three levels of N (5, 15, or 25) with five levels of solution familiarity determined by the rank of the correct item in the normative frequency hierarchy (1, 3, 6, 10, or 15). In each of the 15 conditions, there were 10 *S*s; each *S* received a different order of the 10 problems determined by a Latin square in order to counterbalance specific problems (categories) with ordinal position. The *S*s were 150 introductory psychology students completing a course requirement; in order of their appearance at the laboratory, they were assigned to conditions

---

[1] The assistance of Wendy Rothwell in the collection and analysis of the data is gratefully appreciated.

[2] Requests for reprints should be addressed to Roger L. Dominowski, Department of Psychology, University of Illinois at Chicago Circle, Box 4348, Chicago, Illinois 60680.

TABLE 1

Mean Errors per Problem for All Problems

N	Normative rank of correct item					
	1	3	6	10	15	$\bar{X}$
5	.61	1.98	1.96	1.86	1.68	1.62
15	1.57	5.38	7.06	7.49	5.52	5.40
25	5.02	7.97	10.51	10.85	11.22	9.11
$\bar{X}$	2.40	5.11	6.51	6.73	6.14	5.38

according to a schedule of 10 random orders of the 15 basic conditions.

Each *S* was seen individually and was told that he would be given a series of problems, each of which consisted of a list of names of things belonging to a category, one of which had been selected as correct. Using a sample list, it was explained that the words on a list varied in frequency and that, on every problem, the solution would occupy the same frequency rank (the rank was not specified). Emphasis was given to selecting the correct category member in as few choices as possible. For each problem, *S* was given a list of category members in alphabetical order (list length varying with *N*) and was allowed to keep a record of which words he had selected. Only the list for the current problem was available to *S* at any time.

Results.—For each problem, *S*'s score was the number of errors made prior to selecting the correct item. The analysis of overall performance used each *S*'s mean number of errors per problem for all problems (Table 1). As indicated by the row means, errors increased as a linear function of *N*, $F\ (1, 135) = 427.63$, $p < .001$. The effect of solution familiarity contained both a linear component, $F\ (1, 135) = 75.68$, $p < .001$, and a quadratic component, $F\ (1, 135) = 31.92$, $p < .001$.[3] These effects are depicted by the column means, where it can be seen that errors increased from Rank 1 to Rank 10, then decreased slightly.

The findings of most interest concern the interactive effects of familiarity and *N*. The linear effect of familiarity increased as a linear function of *N*, $F\ (1, 135) = 26.76$, $p < .001$. This reflects the fact that the increase in errors from Rank 1–Rank 15 was greater as *N* increased. The quadratic component of the familiarity effect first increased, then decreased as *N* increased, as indicated by the Quadratic × Quadratic interaction, $F\ (1, 135) = 8.24$, $p < .005$. This interaction can best be described by examining the relative difficulty of problems having the least familiar solutions (Rank 15) for different levels of *N*. With $N = 5$, all problems were quite easy, and selecting the least familiar solution was slightly easier than finding those of intermediate frequency. For $N = 15$, difficulty generally increased, but problems with the least familiar solutions were clearly easier than those with slightly more frequent solutions. However, with $N = 25$, there was both a general increase in errors and a consistent trend toward greater difficulty with decreasing solution familiarity.

There were modest changes in performance over the problem series. Figure 1 depicts the mean number of errors for successive two-trial blocks for each level of solution familiarity. A comparison of the first and last blocks indicated an overall reduction in errors with practice, $F\ (1, 135) = 4.48$, $p < .05$. More importantly, the quadratic component of the familiarity effect increased in magnitude with practice, $F\ (1, 135) = 5.42$, $p < .025$. As indicated in Fig. 1, there was a clear reduction in errors only for Ranks 1 and 15; the latter change resulted in difficulty being noticeably curvilinearly related to solution familiarity at the end of the problem series.

Discussion.—The results can be explained in terms of two strategies for selecting the correct alternative from a set of items forming a frequency hierarchy. One strategy is to guess alternatives in order of decreasing frequency, while the other is to guess alternatives at the "correct" level of familiarity. The fact that errors generally increased with decreasing solution familiarity implies that *S*s try alternatives in order of their frequency, as other studies have shown (e.g., Duncan, 1966). To the extent that *S*s behave in this fashion, adding distractors should prove more detrimental when the solution is less familiar, since more of the distractors will be more familiar than the solution. This expectation was corroborated by the finding that the inverse relationship between errors and solution familiarity was more pronounced as *N* increased.

The alternate strategy of guessing items of appropriate familiarity is related to the curvilinear component of the familiarity effect and its interaction with *N*. The effectiveness of such a strategy depends on the discriminability of the correct alternative (Duncan, 1967). In the present study, with 5 or 15 alternatives, the least familiar solution (Rank 15) was more discriminable than solutions of intermediate familiarity, and fewer errors were made in selecting it (Table 1). However, with 25 alter-

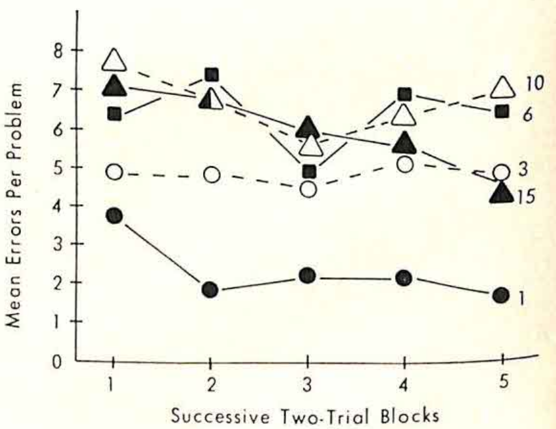


FIG. 1. Effects of practice on finding solutions of different levels of familiarity (rank in a normative frequency hierarchy).

---

[3] Because of the dubious accuracy of the normative ranks and their lack of applicability to problems with five alternatives, the five familiarity levels were treated as an equally spaced ordered series in the analysis.

natives, the least familiar solution no longer had relatively greater discriminability, and finding this alternative was most difficult.

Guessing items of appropriate frequency requires that *S*s have some knowledge of the correct frequency level. In the present study, this knowledge could be acquired during the problem series; consequently, the curvilinear component of the familiarity effect should be most pronounced at the end of the problem series, a result which was obtained. The data of Fig. 1 indicate that guessing items in order of decreasing frequency was the predominant strategy at the beginning of the series, but that *S*s changed to guessing items at the correct frequency level by the end of the series. Supplementary analyses of guessing patterns support this account. On the first problem, *S*s' early guesses were relatively frequent items (in all conditions) and there was a tendency for guesses to be made in order of decreasing frequency. In marked contrast, on the last problem, first guesses regularly decreased in frequency with solution familiarity, and early guesses tended to be confined to appropriate frequency levels. After the problem series, each *S* was asked to estimate the rank (based on frequency) of the solutions; these estimates were reasonably accurate in all conditions. Furthermore, within a condition, the accuracy of these estimates was unrelated to the number of errors made in finding the solutions. Thus, the differential practice effects shown in Fig. 1 reflect differences in the ease with which the correct item could be discriminated from distractors of similar familiarity rather than variation in the accuracy of knowledge concerning the correct frequency level.

An implication of the present findings concerns the effects of problem-solving sets. The adoption of a set, whether through training or instructions, may be described as decreasing the number of alternatives to be considered as possible solutions. If so, these results suggest that an appropriate set should be more effective (relative to a control condition) for less familiar solutions.

REFERENCES

COHEN, B. H., BOUSFIELD, W. A., & WHITMARSH, G. A. Cultural norms for verbal items in 43 categories. (ONR Tech. Rep. No. 22, Contract Nonr-631) Storrs: University of Connecticut, 1957.

DAVIS, G. A. Detrimental effects of distraction, additional response alternatives, and longer response chains in solving switch-light problems. *Journal of Experimental Psychology*, 1967, **73**, 45–55.

DOMINOWSKI, R. L. Problem difficulty as a function of relative frequency of correct responses. *Psychonomic Science*, 1965, **3**, 417–418.

DOMINOWSKI, R. L. Anagram solving as a function of bigram rank and word frequency. *Journal of Experimental Psychology*, 1967, **75**, 299–306.

DUNCAN, C. P. Problem-solving within a verbal response hierarchy. *Psychonomic Science*, 1966, **4**, 147–148.

DUNCAN, C. P. Response hierarchies in problem solving. In C. P. Duncan (Ed.), *Thinking: Current experimental studies*. Philadelphia: Lippincott, 1967.

NEIMARK, E. D., & WAGNER, H. Information-gathering in diagnostic problem solving as a function of number of alternative solutions. *Psychonomic Science*, 1964, **1**, 329–330.

(Received May 19, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 226–228

ROLE OF IMAGERY IN INCIDENTAL LEARNING: REPLICATION AND EXTENSION OF AN EFFECT[1]

PETER W. SHEEHAN[2]

University of New England, Armidale, Australia

The hypothesis was tested that more high-imagery-arousing than low-imagery-arousing noun stimuli will be recalled in incidental (INC) as compared to intentional (INT) learning. The *S*s were tested for recognition of 18 concrete and 18 abstract nouns, controlled for meaningfulness and familiarity. Results supported the prediction and showed that motivation to learn under INT instructions was a determining factor of the strength of the effect.

Imagery defined in terms of the nonverbal responses aroused by concrete stimulus material has been shown to be an effective mediating variable in the recall of material in a variety of memory situations, including tests of free recall and recognition (Paivio, 1969). A recent study (Sheehan, 1971), following an implication of data presented by Sheehan and Neisser (1969), suggested that imagery so defined plays an adaptive role in the incidental learning context. More concrete than abstract noun stimuli were recognized in incidental (INC) as compared to intentional (INT) learning when *S*s inspected stimuli under a familiarity orienting set and were examined for recall by tests of both free recall and recognition, given in that order. Three factors, however, rendered results in the experiment possibly equivocal. First, data may have indicated a chance effect; the differences found were significant only once at the .05 level. Second, the test of recognition was administered after a test of free recall; preceding procedures may have had an important effect on *S*s' recognition performance leading to different results for the two learning groups—clearly, a recognition-alone test is required. Third, only one *E* conducted the study; expectations of *S*s about recall were hence not tested by the most demanding procedure available: an inquiry by a "blind" *E* who has no knowledge of *S*s' learning performance and does not know the identity (INC or INT) of the learning *S*s. The present study aimed to modify the procedures of this experiment specifically to meet these three criticisms and to test for the adaptive role of imagery in INC learning.

Learning relies strongly on the nature of the representational responses available to *S* in the learning situation; and these responses serve to facilitate memory. Imagery is one such representational response and concrete noun stimuli should be recognized equally well in both the INC and INT learning situations since images are easily aroused by such stimuli at inspection. As the imagery-arousing value of the noun stimuli diminishes, however, more effective representational responses (such as verbal coding) should be aroused in the intentional context since INT *S*s are motivated by the knowledge that they later will be asked to remember and INC *S*s do not have any special set to learn. Memory facilitation by imagery rather than by verbal means also should be more likely to occur in the unexpected recall situation than in INT learning, for in the latter context nonverbal and verbal mediators are both normally freely available to aid memory. Thus there should be no difference or a small difference between intentional and incidental learners in the recognition of concrete stimuli, but a relatively large difference in the recognition of abstract stimuli. Hence, the hypothesis of this study follows that there will be a larger difference in recognition accuracy between concrete and abstract noun stimuli under incidental than under intentional learning conditions. The incidental learning context is defined, here, as one where *S*s clearly do not expect they will be later asked to recall.

Method.—Independent sets of 20 (10 males and 10 females) first-year psychology students were allocated to each of four learning conditions: INC, INT, INT-Motivated, and INT-Control. All 80 *S*s were unpaid volunteers. The INC and INT conditions essentially replicated the confirming conditions of the previous experiment, but tested for the predicted effect more adequately by meeting the objections outlined above; the remaining two conditions provided new information about the predicted effect.

Stimuli were 18 concrete (I) and 18 abstract (N-I) nouns printed in matt-black ink on 5 × 3 in. index cards. Each noun was exposed tachistoscopically for 4 sec. The imagery index values for the concrete (e.g. MONK, NAIL) and abstract (e.g., CHANCE, THEORY) nouns were 6.25 and 3.08, respectively, and words were equated for both meaningfulness, *m*, and familiarity, *f*. Word values were selected from the list as published by Paivio, Yuille, and Madigan (1968). Two variations of noun ordering were used at inspection: random arrangement of I and N-I stimuli, and random arrangement reversed. Within each experimental group *S*s were alternated from one list to the other. All *S*s were given two practice stimuli (TABLESPOON and SENTIMENT) to familiarize them with experimental procedures.

The "nonlearning" nature of the study was implied to INC *S*s by initial instruction to all

---

[1] This research was supported in part by a grant from the Australian Research Grants Committee. The author is indebted to Jacqueline Lloyd for her assistance as E_1 in the experiment.

[2] Requests for reprints should be sent to Peter W. Sheehan, Department of Psychology, University of New England, Armidale, New South Wales, Australia 2351.

groups, after *S*s had entered the experimental room, that *E* was interested in the measurement of individual differences in rating judgments on a number of different variables. The *S*s in the INC, INT, and INT-Motivated groups were asked to judge how familiar they were with each noun stimulus. In each of the three groups, *S*s rated the familiarity of individual stimulus items on a 7-point rating scale measuring rated frequency of word usage in their daily speech. Ratings were recorded by *S*s in score booklets, and the time taken by them to make their judgments was also recorded, as measured by a chronoscope timer which was started automatically at stimulus onset and stopped by *S* pressing a key to indicate he had made the required rating. The INT-Control *S*s did not receive the above-mentioned familiarity orienting task at inspection. After initial instruction they were simply asked to press the key to indicate they had seen each stimulus item.

"Intent to learn" was manipulated by means of verbal instruction. The *S*s in the INC condition received no instruction prior to stimulus inspection other than *E*'s request for them to judge the familiarity of the stimuli. The *S*s in the INT condition were treated identically except, in addition, they were told, just prior to receiving the practice stimuli, that they would later be asked to recall as many words as they could remember. The *S*s in the INT-Motivated group were instructed at the same point that not only would they be later asked to recall but *E* expected them to try hard to learn as many of the words as they could. The *S*s in the INT-Control condition were instructed to learn the stimulus in an identical fashion to that used for the INT-Motivated group. Instructions and procedures for the four groups were otherwise similar. Practice and experimental stimuli were presented once, and 30 sec. after the presentation of the last stimulus all *S*s were tested for recognition. The *S*s were shown a list of 100 nouns containing the 36 experimental stimuli and 64 additional "distractor" nouns varying in imagery-arousing value. All *S*s circled the nouns they recognized on answer sheets provided. No time limit was set for their task.

Two *E*s were involved in the study. The E_1 tested learning in the session proper and inquired following the recognition test into the nature of *S*s' imagery at inspection and at the moment of recall. The *S*s were then taken to E_2 who conducted a comprehensive inquiry into their expectations about the possibility of recall. The second *E* was "blind" both as to the identity of each experimental *S* and to the nature of *S*s' actual recognition performance; his inquiry ranged from general to very specific probes into *S*s' perceptions of the purpose of the study. The "blind" procedure eliminated 14.3% of the *S* sample, whereupon additional *S*s were tested to equate experimental groups in *S* size ($n = 20$).

TABLE 1

Mean Reaction Times for Ratings Given to Concrete (I) and Abstract (N-I) Stimuli for Subjects Tested in Each of the Four Experimental Groups

Experimental group	Stimulus imagery value	
	I	N-I
INC	2.03	2.20
INT	2.47	2.66
INT-Motivated	2.38	2.50
INT-Control	2.35	2.42

Note.—$n = 20$ per group.

TABLE 2

Mean Number of I and N-I Stimuli Recognized by Male and Female Subjects in Each of the Four Experimental Groups

Experimental group	Stimulus imagery value	
	I	N-I
INC		
Male	16.9	12.2
Female	16.4	12.0
INT		
Male	16.7	13.6
Female	16.9	13.6
INT-Motivated		
Male	16.9	14.6
Female	17.6	15.7
INT-Control		
Male	14.8	12.9
Female	16.1	14.3

Note.—$n = 10$ per group.

Results and discussion.—Mean reaction times for key pressing given to the two kinds of imagery arousing stimuli by *S*s in the four experimental groups are listed in Table 1. Reaction time data were analyzed by a 4 × 2 (Experimental Group × Stimulus Imagery Value) split-plot analysis of variance. Analysis showed a significant main effect for stimulus imagery value; over all groups, *S*s took longer to respond at inspection to N-I stimuli than they did to I stimuli, F (1, 76) = 17.13, $p < .001$. Interaction analysis, however, showed that the INC, INT, INT-Motivated and INT-Control conditions were equivalent in the time discrepancy between the two types of stimuli, $F < 1$. Results indicated that although *S*s may have taken longer to respond to the abstract than they did to the concrete stimuli, the differences in reaction time to the two kinds of stimuli did not vary among the separate groups. Hence, observed differences in the accuracy of recognition of I as compared to N-I stimuli from one group to the other cannot be attributed to group-specific variation in response time at inspection.

Table 2 sets out the data relevant to accuracy of recognition and lists the average number of I and N-I stimuli correctly recognized by each of the four experimental groups. Scores were subdivided to take account of the sex of *S*s and a 4 × 2 × 2 (Experimental Group × Stimulus Imagery Value × Sex of *S*) split-plot analysis of variance conducted to analyze the differences among *S*s in their accuracy of recognition. Analysis indicated significant main effects for experimental group, F (3, 72) = 3.41, $p < .05$, and stimulus imagery value, F (1, 72)

$= 100.36$, $p < .001$; and a significant interaction effect between stimulus imagery value and experimental group, $F\ (3,\ 72) = 4.45$, $p < .01$. The interaction test result directly supports the prediction under test that the difference in recognition between concrete and abstract noun stimuli would be greater in incidental than in intentional learning. Individual t-test comparisons across sexes of the mean differences among learning groups indicated that the INC condition differed appreciably from all intentional learning conditions: the INC group differed from the INT condition, $t\ (72) = 1.64$, $p < .06$, the INT-Motivated condition, $t\ (72) = 2.97$, $p < .005$, and the INT-Control condition, $t\ (72) = 3.27$, $p < .005$. Results suggest that earlier found differences between the INC and INT conditions under familiarity orienting set (Sheehan, 1971) were small but real. Data also extend previous findings. They show that the predicted effect is strengthened considerably when the learning nature of the intentional condition is emphasized to *S*s (as in the INT-Motivated and INT-Control conditions) by *E* asking *S*s to make an effort to learn the material that is presented to them.

Close analysis of the data indicated that there were small differences between intentional and incidental learners in the recognition of concrete stimuli, but relatively large differences in the recognition of the abstract words. Comparing the INC and INT-Motivated conditions, for example, for recall patterns of I and N-I stimuli (considered separately) evidence, showed that I recognition tended to be maintained under the expected and unexpected recall conditions, $t\ (38) = 1.24$, $p > .10$, while N-I stimuli were much more poorly recognized under unintentional than under intentional instruction, $t\ (38) = 3.33$, $p < .01$. Further analysis showed that it was the number of correct stimuli recognized that was pertinent to the predicted effect; the four experimental conditions did not differ ($p > .05$) in the number of false I or N-I positives recognized. In addition, data from E_1's inquiry into the nature of *S*s' cognitive processes when stimuli were first presented showed no differences among the four groups in the vividness of mental imagery (as measured on a 7-point rating scale) reported by *S*s to be present at inspection, $F\ (3,\ 76) = 1.52$, $p > .10$. This result indicates that nonverbal reactions, viewed as representational responses elicited by stimuli at presentation, were equally available to facilitate recognition for the incidental and the intentional learning groups. Since stimuli adopted for the study were controlled for meaningfulness and familiarity, the obtained variation in I and N-I patterns of recognition performance seems better explained in terms of the varying capacity of concrete and abstract noun stimuli to arouse sensory imagery than by appeal to verbal mediational strategies.

Table 1 shows clearly that the pattern of differences in I/N-I recognition among the various groups is virtually identical for both males and females. There was no evidence of appreciable sex differences in the overall accuracy of recognition for either type of imagery stimulus; and males and females showed equal amounts of INC learning, $t\ (18) = .39$, $p > .10$. These results conflicted to some extent with the findings of Ernest and Paivio (1971) who reported significantly higher INC learning for females than for males on an item-recognition task.

The emergence of a similar-sized effect for independent *E*s in separate studies attests to the consistency of the phenomenon being investigated. Results suggest that previous findings did not reflect chance outcomes, or effects due simply to the contamination of the recognition measure by *S*s' previous recall of test stimuli. Considering the replicability of test results for recognition and the initial absence of the predicted effect in free recall (Sheehan, 1971), further work needs to explore the extent to which imagery effects in INC learning are relevant to particular tests of memory.

REFERENCES

Ernest, C. H., & Paivio, A. Imagery and sex differences in incidental recall. *British Journal of Psychology*, 1971, **62**, 67–72.

Paivio, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, **76**, 241–263.

Paivio, A., Yuille, J. C., & Madigan, S. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 1968, **76**(1, Pt. 2).

Sheehan, P. W. The role of imagery in incidental learning. *British Journal of Psychology*, 1971, **62**, 235–243.

Sheehan, P. W., & Neisser, U. Some variables affecting the vividness of imagery in recall. *British Journal of Psychology*, 1969, **60**, 71–80.

(Received June 1, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 229–230

ENCODING VARIABILITY AND IMAGERY: EVIDENCE FOR A SPACING-TYPE EFFECT WITHOUT SPACING[1]

DONALD ROBBINS[2] AND PHILLIP S. WISE
Emory University

Thirty *S*s were tested for recall of medium-imagery noun pairs after one or two presentations. The presentations were separated by a high- or low-imagery noun pair, which was assumed to be an easy or difficult intervening task, respectively. After one presentation, fewer errors occurred following an easy intervening task, while after two presentations the reverse was found. The results were interpreted as supporting an encoding variability view of retention.

This article deals with an encoding variability view of retention. Essentially this view states that when an item is presented, *S* stores a coded version, and subsequent repetitions provide opportunities for additional codes to be stored. The more variable the codes, the more likely *S* is to recall the item.

Recently, Bjork and Allen (1970) devised an ingenious method for testing an encoding variability notion. Using a modified Brown-Peterson paradigm, they varied the processing difficulty of the task that intervened between two presentations of an item. The results indicated that after a single presentation, items followed by an easy intervening task were recalled better than items followed by a difficult intervening task. However, on the retention test following a second presentation (which had a task of moderate difficulty following it), the difficult intervening task condition showed as much recall or more than the easy intervening task condition. Bjork and Allen concluded that their data supported an encoding variability theory on the assumption that the difficult intervening task leads to more variable codes than the easy intervening task.

In the present study, another method was used to manipulate the difficulty of the intervening task. Paivio (1969), among others, has shown that there is a direct relationship between ease of visual imagery and recall of pairs of English words. Specifically, he and his colleagues (see Paivio, 1969) have shown that pairs of high-imagery words are better recalled than pairs of low-imagery words. In the present study, a continuous, paired-associate, study-test procedure was used, where the critical test items were medium-imagery pairs. The critical items that were presented twice were separated by either a high- or low-imagery pair, which were assumed to be an easy or difficult task, respectively. In addition, some items were presented only once where the presentation and the test were separated by either a high- or low-imagery pair. It was expected that after a single presentation the easy intervening pair condition would show fewer errors on the test than the difficult intervening pair condition; while for the repetition conditions, the difficult intervening pair condition should show the same number of or fewer errors than the easy intervening pair condition.

Method.—The critical items were pairs of medium-imagery nouns which were given one or two presentations. For the single-presentation conditions, the presentation of a medium-imagery pair was followed by either a high-imagery pair (which was assumed to be an easy task) or a low-imagery pair (which was assumed to be a difficult task), and then by a test of the medium pair. For the repetition conditions the two presentations of a medium-imagery pair were separated by a high-imagery or low-imagery pair, followed by two presentations of two different medium-imagery pairs and then a test of the repeated medium pair. In order to prevent *S*s from detecting the structure of the sequence, some repetition conditions were constructed in which a single rather than a repeated item was tested. As a result, each *S* had 12 test trials of the easy intervening and 12 test trials of the difficult intervening single-presentation conditions and 6 test trials of the easy and 6 of the difficult intervening repetition conditions. As an additional precaution, single-item-presentation trials were administered which were followed by an immediate test—the tested items in these cases were not the first pair presented, a medium-imagery pair, but rather either high-imagery or low-imagery pairs. Each *S* had six such pairs. The conditions were presented to *S* in a randomized block order. In addition the first six pairs were not tested.

Thirty undergraduate students from an introductory psychology course at Emory University fulfilled part of a course requirement by participating in the study.

The stimulus materials were pairs of low-low, medium-medium, and high-high imagery nouns taken from the Paivio, Yuille, and Madigan (1968) norms, with means of 2.52, 4.75, and 6.65 for low-, medium-, and high-imagery nouns, respectively. The word pairs were printed on 3 × 5 in. index cards for study trials, and the left-hand member of each pair was printed on the left-hand side of a 3 × 5 in. index card for test trials. Three different lists were compiled and two different orders (of the various conditions) within each list yielded six sublists. Five *S*s were tested on each sublist. An interval timer made a loud audible click every 6 sec., which was a signal to turn to the next card. Both study and test trials were of 6-sec. duration.

---

[1] This research was supported in part by a grant from Emory University, Graduate School.

[2] Requests for reprints should be sent to Donald Robbins, Department of Psychology, Emory University, Atlanta, Georgia 30322.

An experimental session, which lasted approximately 30–35 min., consisted of 222 trials (174 presentation trials and 48 test trials). The *S*s were tested individually and given interactive visual imagery instructions, and the nature of study and test trials were explained to them. On study trials they were to read the words aloud and then form the image; on test trials they were to read the word aloud and then give the response vocally. No information was given on test trials. The cards were placed on a table and presented face up to *S*, one at a time. Since it took approximately .5 sec. to turn a card, the presentation rate was approximately one card every 6 sec. Thus, for the single-presentation conditions, the retention interval was approximately 6 sec., except for the immediate tests where it was approximately .5 sec. For the repetition conditions, the interval between the repetitions was approximately 6 sec. while the retention interval was approximately 12 sec. (two items).

Results and discussion.—Since there were no obvious differences between sublists, the data for all 30 *S*s were combined. A lenient scoring procedure was used and since there was only one case where a word which was recalled was not recalled in the exact form at presentation, a lenient or strict scoring procedure yields almost identical results.

The mean percent errors for each condition was calculated and revealed that after one presentation, after an easy intervening pair, *S*s made 25.6% errors; while after a difficult intervening pair, *S*s made 35.3% errors. However, after two presentations, 23.9% in contrast to 35.6% errors were made for the difficult versus the easy intervening pair conditions. A Sandler's *A* test (McGuigan, 1960) was performed on the difference between the number of errors in the easy and difficult condition for each *S* and revealed a significant difference for the one-presentation condition, $A\ (29) = .195$, and for the repetition conditions, $A\ (29) = .202$; $p < .05$ in both cases, since $A\ (29) \leq .264$ for the 5% level of significance. In addition, *S*s made 2.2% errors on the immediate tests for both high- and low-imagery pairs.

The results of the present study reveal that after a difficult intervening pair, *S*s made more errors in recall for the prior item than after an easy intervening pair. When a second presentation was given, rather than a test, a subsequent recall test revealed a reversal: those pairs whose repetitions were separated by a difficult pair were recalled better than those pairs whose repetitions were separated by an easy pair.

The findings that *S*s made more errors after two presentations than after one presentation in the easy intervening task condition may be the result of a longer retention interval, different numbers of intervening pairs or a combination of the two, if one assumes that minimal processing occurred on the second presentation of the pair.

Thus, these results extend those of Bjork and Allen (1970), who using a different paradigm found a similar reversal, and thus may be viewed as further support for an encoding variability interpretation of retention.

REFERENCES

BJORK, R. A., & ALLEN, T. W. The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 567–572.

MCGUIGAN, F. J. *Experimental psychology.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

PAIVIO, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, **76**, 241–263.

PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 1968, **76**(1, Pt. 2).

(Received June 5, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 231–233

MEMORY SPAN: EFFECTS OF STRING LENGTH AND STRING COMPOSITION

BAYLA MILLER MYER AND DANIEL C. O'CONNELL[1]
Saint Louis University

Memory span was tested for symbol and digit strings. Symbol strings were sequentially exposed on a memory drum for 3 sec., and digit strings for 1.5 sec.; *S*s recorded recall when the exposure interval elapsed. Memory span of symbol strings decreased as string length or number of different symbols composing the strings increased. Recall of digit strings was a function of string length only. Well-established storage and retrieval strategies (viz., grouping) are evidently available to increase digit recall.

Miller (1956) asserted a fixed limit of about seven units or chunks to man's short-term memory. Memory is incrementally extended by organizing multiple information within this limited capacity. Three relevant factors that concern memory span of strings are (*a*) length of the string to be recalled, (*b*) *S*'s familiarity with the material that composes the string, and (*c*) grouping of the elements within a string.

Crannel and Parrish (1957) found that grouping techniques increased the recall of digit sequences, but that difficulty of string recall was not affected as number of different elements composing strings increased; they did not vary the conditions systematically. Williams and Fish (1965) systematically varied string length and number of different elements composing digit and symbol strings to assess immediate retention differences between familiar and unfamiliar strings of different lengths and compositions. Short-term retention of strings decreased as string length, number of different elements within strings, and material unfamiliarity increased. These results contrasted directly with the finding of Crannel and Parrish that digit-string recall difficulty was not related to composition. The present study replicated, extended, and reinterpreted Williams and Fish.

Method.—Six *S*s were tested in each condition in a repeated-measures design. There were 36 stimulus conditions representing all possible combinations of string lengths (2–9 total elements) and number of different elements composing the strings (2–9). Stimulus conditions formed a half matrix as in Table 1.

Nine experimental symbols were used: asterisk, circle, diamond, cross, triangle, slash, semicircle, square, and upside-down T. Symbol height was $\frac{5}{16}$ in. The symbols were all easily verbalized and were all equally recognizable (Williams & Falzon, 1963). Each symbol was centered on a $\frac{3}{8}$-in. square white background and sequences were affixed to the memory-drum paper with nonreflective Scotch Magic tape. Elements for digit strings were numbers 1–9 in standard pica type. Single blank typewriter spaces were between adjacent elements in digit-string construction. The *S*s received two sets of symbol strings and two sets of digit strings. Each set contained 36 strings representing all possible combinations of string lengths and compositions. String composition and sequences of strings within lists were determined randomly without replacement. A postexperimental questionnaire was given concerning ease of string recall, strategies for retention, and material differences.

The *S*s were seated 15 in. from a Stowe memory drum window. Instructions for ordered recall were read by each *S*, and each *S* received information concerning the specific sequence in which the four lists were to appear; list sequence orders were balanced among *S*s. Five practice trials were given before each actual experimental list to familiarize *S* with the stimulus materials. Strings were presented sequentially one at a time. Symbol list strings were exposed for 3 sec. with 12 sec. for recall, digit strings for 1.5 sec. with 6 sec. for recall. A previous

TABLE 1

MEAN PERCENTAGE RECALL SCORES FOR SYMBOL (UPPER ENTRIES) AND DIGIT (LOWER ENTRIES) STRINGS

String length	Number of different symbols or digits in string							
	2	3	4	5	6	7	8	9
2	100.0							
	100.0							
3	100.0	100.0						
	100.0	100.0						
4	97.9	100.0	93.8					
	100.0	100.0	100.0					
5	96.5	93.4	81.6	68.3				
	100.0	100.0	97.8	96.7				
6	82.0	80.6	61.2	55.6	43.1			
	98.5	93.0	94.4	90.3	97.2			
7	85.8	67.9	45.3	59.6	44.0	40.5		
	95.3	84.6	84.0	86.9	90.5	93.9		
8	77.1	63.5	37.5	50.0	39.6	39.6	41.7	
	78.2	75.0	75.0	82.4	78.1	82.3	81.3	
9	86.2	39.8	42.5	28.8	34.3	24.1	34.3	31.5
	78.7	78.6	61.5	69.5	73.2	68.5	67.6	65.8

---

[1] Requests for reprints should be sent to Daniel C. O'Connell, S. J., Department of Psychology, Saint Louis University, 221 North Grand Boulevard, Saint Louis, Missouri 63103.

investigation (Williams & Fish, 1965) indicated that these exposure rates were equivalent for difficulty of recall. The *E* observed *S* in a mirror to make sure that *S* did not begin written recall before the string disappeared from view. Each *S* completed two sets and had a 5-min. break before finishing the remaining two sets.

Six males voluntarily served as *S*s. They were psychology graduate students at St. Louis University and were not necessarily naive to memory studies.

Results.—Due to the incomplete matrix representing the stimulus conditions, only four string lengths (6–9) and five composition conditions (2–6) were included in the analyses of variance for the digit and symbol data. Two lists of each string type (digit and symbol) were used; therefore a trials factor was included to determine practice or fatigue effects. The analysis of variance model is String Length × Composition × Trials.

The dependent variable was percentage of correct digits or symbols correctly positioned within each string. Mean recall scores are given in Table 1. Symbol string recall decreased as either string length or the number of different symbols composing the string increased; digit-string recall decreased as string length increased, but remained consistent as different compositions varied.

Variance analyses for percentage data required an arc-sine transformation. A conservative *F* test ($df = 1, n$) was used to determine statistical significance since a repeated-measures design was employed. In the symbol analysis there were two significant main effects, string length, $F(1, 4) = 7.03$, $p < .05$, and number of different elements composing the string, $F(1, 5) = 23.63$, $p < .005$. No interaction effects were significant. The absence of practice or fatigue effects was indicated by nonsignificance of the trial factor.

Analysis of digit-string data yielded a significant *F* for string length, $F(1, 4) = 30.44$, $p < .005$; the main effect of number of different elements composing the string was not significant, $F(1, 5) = 1.33$. There were no interaction or trial effects.

Discussion.—Symbol recall decreases as either string length or number of different symbols composing strings increases. Symbols are not easily grouped to facilitate storage and retrieval; each symbol and its location in the string must be separately encoded. Grouping is effective to increase recall (Crannel & Parrish, 1957), but grouping techniques are unavailable to *S*s for sequences of symbols. Although the symbols used are common, everyday usage does not require memory of sequences; *S*s do not have effective chunking strategies and the memory limit is quickly reached.

Recall of digit strings was affected by length alone, not by composition: encoding digit sequences was so effective that equal recall of string length was obtained whether there were few or many different digits composing strings. Since digits are familiar, *S*s are used to storing and retrieving digit sequences and have effective memory strategies available. Grouping digits facilitates recall and reduces the separate cues necessary for encoding. Digit-string recall was superior to symbol recall even though exposure time was half as long; familiarity and chunking differences can account for this result.

These findings agree with Williams and Fish (1965) with two exceptions. Williams and Fish found that increasing the number of different elements composing the strings decreased the recall for both symbol and digit materials. The present study does not find support for a significant effect of the number of different digits on recall of digit strings, but agrees with the finding of Crannel and Parrish (1957) on digit-string recall. The Williams and Fish variance analysis on symbol data yielded a significant Length × Composition × Trials interaction which the authors could not explain; there was no indication of this three-factor interaction in the present study.

The questionnaire data supported the interpretation of recall differences offered above. The *S*s divided strings into "sensible parts" and "patterns" and agreed they were unable to use the same strategy for symbols and digits because digit strings were easier to encode. Groups of digits were stored, whereas symbols were learned separately.

There is empirical evidence suggesting that the strategy used for digit retention may be inhibiting rather than facilitating when applied to symbols. Spacing of repeated elements within strings affects the strategy *S* employs. When repeated elements are well spaced or occur at the ends of the string, differentiation among elements and retention are facilitated. Inspection of the idiosyncratic strings composing symbol lists indicated that the Ranschburg Effect (RE) accounts for much variability. RE (Jahnke, 1969) is an inhibitory phenomenon in memory that occurs when a string contains a repeated element; the repeated element tends to be incorrectly recalled because *S* is not sure of the number of repetitions or the number of elements that are repeated. This inhibition occurs readily when strings are easily perceived as containing repetitions. When repeated symbols were spatially close, the use of grouping could have inhibited retention.

Since the questionnaire indicated that digits were stored by internal verbal organization, whereas symbols were stored by perceptual mechanisms, a post hoc inspection of symbol data was conducted. Since symbols were not equated for total occurrences in strings, distribution of intrusion errors for each symbol was the concern, not comparisons of total incidences of intrusion errors. Errors were not randomly distributed over all symbols. The data suggest that errors involved perceptual processes: *S*s recorded crosses as the modal error (9 out of 41 total errors) when an asterisk was required (which contains the cross); circles were misrepresented most frequently as semicircles (10 out of 40), a component of both symbols; diamond errors were mostly accounted for by triangles (52 out of 102), and vice versa (23 out of 78); and upside-down T and slash confusions (both symbols being perceptual components of the asterisk) accounted for 41 of the 98 total errors made involving these two symbols.

REFERENCES

CRANNEL, C. W., & PARRISH, J. M. A comparison of immediate memory for digits, letters, and words. *Journal of Psychology*, 1957, **44**, 319–327.

JAHNKE, J. C. The Ranschburg Effect. *Psychological Review*, 1969, **76**, 592–605.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **63**, 81–97.

WILLIAMS, J. R., & FALZON, R. P. The relationship of display system variables to symbol recognition and search time. *Journal of Engineering Psychology*, 1963, **2**, 97–111.

WILLIAMS, J. R., & FISH, D. L. Effect of item length and number of different elements on immediate memory. *Psychonomic Science*, 1965, **3**, 353–354.

(Received June 17, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 233–234

VISUAL SERIAL SEARCH PERFORMANCE FOR NUMBER AND LETTER TARGETS[1]

S. VITERBO McCARTHY[2]

Regis College, Massachusetts

To discover whether visual serial search time for letters differed significantly from visual serial search time for numbers, a repeated-measurement design was used with two groups of 20 undergraduate female *S*s. Results established the superiority of visual serial search performance for number targets over letter targets when 60 sec. or 0 sec. intervened between 10, 30-sec. visual serial search trials on variable forms. The observed effect was attributed to the higher frequency of occurrence of number target and number field components.

Typically, number targets are used for studying visual serial search performance (Ammons, 1955; Underwood, 1961; Weber, Carlton, & Perry, 1970). However, the question has been raised by Underwood (1966): "will search time for letters be different from search time for numbers [p. 285]?"

A comparison of the critical features of the 26 capitalized alphabetic targets with the critical features of the 26 arabic number targets 0–25 suggests that all components of the alphabetic targets are distinctive, whereas there are only 10 distinctive components or couplings thereof for the numeric targets. Thus, a serial search for 26 consecutive alphabetic targets would assure seeing each component once, while a serial search for the same number of consecutive numeric targets would probably mean seeing each component more than once. The number of possible repetitions would range from 2 to 13 depending upon the component and its location within the target number; over several trials, repetitions or frequency of occurrence would become still more marked. That more frequently occurring targets are searched more rapidly has been demonstrated for *discrete* visual search tasks (Krueger, 1970). If Krueger's findings also pertain to serial visual search tasks, then visual serial search time for number targets should be significantly faster than visual serial search for letter targets.

Method.—To test this hypothesis, a repeated-measurements design was used: each *S* searched both number and letter targets over 10 practice trials with either 0-sec. or 60-sec. intertrial intervals. Across number and letter fields, forms were matched for step size (distance between consecutive targets) and variability of step size (a changing or fixed distance between consecutive targets), constraints *within* any single trial which affect visual serial search performance (Weber, Carlton, & Perry, 1970). Five forms were used with both number and letter targets; each was repeated once after five trials; there was an incomplete between-*S*s counterbalancing of forms across trials. For either 0-sec. or 60-sec. intertrial intervals, variable forms which are less affected than constant forms by specific learning were used, since constraints *across* visual serial search trials which are known to alter performance include intertrial intervals and constant or variable forms (Ammons, 1955; Underwood, 1961).

Each variable form had 26 letter targets, A to Z (L Forms), and 26 number targets, 0 to 25 (N Forms). Letter target "A" or number target "0" was centered and underlined on each page of L Forms or N Forms, respectively. The targets were sprinkled over an 80 × 80 space covering 5 in.^{2} on $8\frac{1}{2} \times 11$ in. paper. To position targets in this coordinate space, 25 ordered pairs were generated from a table of random number. Each letter from B to Z was randomly assigned an ordered pair which described its position on the serial search form; this procedure was repeated to obtain five variable L Forms. The 26 number targets on the five variable N Forms were positioned so as to correspond with

---

[1] Research support was provided by Grant GY-7563 from the National Science Foundation to Regis College. The author wishes to thank Linda Faldetta and Jane Grenier for their assistance in the conduct of the experiment, and Warren D. Gribbons and Elizabeth Shuhany and their students for cooperation in this research.

[2] Requests for reprints should be sent to S. Viterbo McCarthy, Department of Psychology, Regis College, Weston, Massachusetts 02193.

TABLE 1

Mean and Standard Deviation for Targets Searched over Ten Trials for Twenty Subjects

Trial no.	Cond.							
	60-sec. intertrial interval				0-sec. intertrial interval			
	N Form		L Form		N Form		L Form	
	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*
1	13.80	2.82	12.85	4.87	15.30	3.91	11.45	4.71
2	15.90	3.48	13.30	4.27	16.75	2.90	12.35	3.29
3	16.55	3.69	15.15	3.88	15.65	4.67	12.85	2.61
4	17.85	3.58	15.05	3.93	15.40	2.50	13.75	3.75
5	17.45	4.18	14.85	3.81	16.05	2.69	13.70	3.07
6	18.40	3.32	15.55	3.62	16.50	3.90	13.40	3.80
7	18.75	4.55	16.05	4.36	16.80	4.27	15.00	3.38
8	18.60	3.99	17.70	4.10	17.45	4.44	15.35	3.04
9	18.35	3.26	16.55	3.85	17.20	3.22	15.00	4.10
10	18.70	4.84	16.95	4.61	16.35	3.93	14.60	2.90
Overall	17.43	3.77	15.40	4.13	16.34	3.64	13.74	3.46

the 26 letter targets as they occurred on the five variable L Forms.

Two groups of 20 undergraduate female psychology concentrators from Regis College, Weston, Massachusetts served as *S*s. All participated as part of a course requirement.

The testing sessions for the two groups took place in a large experimental psychology laboratory. At each session, two booklets containing 10 appropriately counterbalanced L Forms and 10 appropriately counterbalanced N Forms were distributed to each of 20 *S*s. According to seating position, alternate *S*s worked on the L Forms first and then on the N Forms; the remaining *S*s worked in the reverse order. Two *E*s were used to monitor *S*s and to time trials. Across trials, *S*s were directed to search and connect as many letters in alphabetical order or numbers in consecutive order as possible. Commands to "Start" and "Stop" signaled the beginning and end of each trial. For each *S*, the number of targets connected on each trial was recorded.

For Group 1, half of the *S*s completed 10, 30-sec. visual serial search trials through randomly ordered letter fields followed by 10, 30-sec. visual serial search trials through randomly ordered number fields with a 60-sec. rest interval between all trials. For the remaining half of the *S*s, the same procedure was followed, but the order of booklets was reversed so that number targets were searched for the first 10 trials and letter targets for the last 10 trials. For Group 2, a similar procedure was followed except that no rest interval occurred between trials.

Results and discussion.—A three-way analysis of variance (*S*s × Trials × Targets) was carried out for the number of targets connected by each of 20 *S*s across 10 trials on Forms L and N. For the 60-sec. intertrial intervals, $F\ (1, 19) = 15.67$, $p < .01$; with 0 sec. between trials, $F\ (1, 19) = 24.90$, $p < .0001$.

Under distributed conditions, number targets were searched significantly faster than letter targets for the population sampled; these results were replicated with Group 2 when no rest intervened between trials. Moreover, there was a *consistently* higher performance with number targets than with letter targets on successive trials following both 60-sec. and 0-sec. intertrial intervals. The mean responses over 10 trials for 20 *S*s on Forms L and N for both groups are listed in Table 1.

One other main effect reached significance on this mixed model: across trials with 60 sec. intervening, $F\ (9, 171) = 9.33$, $p < .001$; with no intertrial rest over trials, $F\ (9, 171) = 3.42$, $p < .01$. The significant improvement which occurred across trials is in line with earlier results obtained by Ammons (1955) and Underwood (1961).

In response to the question raised by Underwood (1966), the results established the superiority of visual serial search performance for number targets over letter targets when either 60 sec. or 0 sec. intervened between 10, 30-sec. visual serial search trials on variable forms. A partial explanation of the observed effect may be the higher frequency of number target components than alphabetic target components; this would justify an extension of Krueger's (1970) findings to include serial visual search tasks.

The type of target field may also account for the observed differences: when Krueger (1970) examined discrete visual search tasks for nontargets of high, medium, and low frequency, search time was significantly shorter for the most frequent nontargets. By analogy, in serial visual search across alphabetic and numeric fields, numbers and letters which are not being searched as targets at any particular instant may be considered nontargets which are being scanned though possibly not seen. During the scanning process, components of number fields with frequencies of 13, 8, 3, 2 would be scanned more often than alphabetic components of letter fields with a frequency of 1. The prediction would follow that the scanning of certain high-frequency components in number fields (nontargets) would contribute to a shorter search time than the scanning of the unit frequency components in letter fields (nontargets).

REFERENCES

Ammons, C. H. Task for the study of perceptual learning and performance variables. *Perceptual and Motor Skills*, 1955, **5**, 11–14.

Krueger, L. E. Effect of stimulus probability on two-choice reaction time. *Journal of Experimental Psychology*, 1970, **84**, 377–379.

Underwood, B. J. Distributed practice on the Tsai-Partington Number Test. *Perceptual and Motor Skills*, 1961, **12**, 325–326.

Underwood, B. J. *Experimental psychology*. New York: Appleton-Century-Crofts, 1966.

Weber, R. J., Carlton, P., & Perry, W. Visual search through random walk number fields. *Psychonomic Science*, 1970, **18**, 207–209.

(Received July 2, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 235–237

ESTIMATES OF CONDITIONAL PROBABILITIES OF CONFIRMING VERSUS DISCONFIRMING EVENTS AS A FUNCTION OF INFERENCE SITUATION AND PRIOR EVIDENCE[1]

PHILIP BRICKMAN[2] AND SCOTT M. PIERCE

Northwestern University

Negative recency effects were predicted and obtained in revisions of the subjective conditional probability of data given a hypothesis. Furthermore, greater effects were indicated for real-life inference situations (judgments of an international crisis, a group, or a person) than for an abstract situation (judgments in a poker-chip problem). Implications are discussed for the phenomena of conservatism and primacy effects in Bayesian inference and also for the notion of representativeness in a sample.

In most previous work involving Bayesian inference (e.g., Edwards, 1968), the conditional probability of observing an event given a particular hypothesis, or P(D/H), is fixed by *E* at the beginning and is assumed to remain constant. However, the well-known phenomenon of gambler's fallacy (cf. Jones, 1971), whereby *S*s expect a change in outcome following a run of similar outcomes, suggests that the probability of another similar event will decrease as past observations of that event increase. According to Bayes's theorem, accumulation of confidence in a hypothesis should be proportional to both P(H), the prior probability of the hypothesis, and P(D/H). If we find that P(D/H) decreases as P(H) increases, we may have evidence for an additional mechanism contributing to the conservatism displayed by *S*s in their accumulation of confidence in a hypothesis.

The present study gathers direct evidence on what happens to P(D/H), subjectively, as evidence accumulates. These effects are compared in four different inference situations: an abstract poker-chip problem and three simulated real-life problems (involving Russia's intentions in a military situation, the attractiveness of a group, and the likeability of an individual, respectively). Since previous work (e.g., Alker & Hermann, 1971) has demonstrated that conservatism effects are greater in "important" real-life decision situations, we might expect stronger effects on P(D/H) estimates in these situations.

The present study also manipulated whether or not *S*s were specifically instructed to regard the events in the sampled sequence as independent or nonindependent. If *S*s believed that the events being observed were correlated for reasons irrelevant to the truth of the hypothesis in question (i.e., because of correlated error), they should then be even more reluctant to accumulate confidence in a hypothesis following a string of favorable events than if they believed that these events were independent.

Method.—The *S*s were 80 Northwestern University undergraduates who earned credit in their introductory psychology course for participating in the experiment. Ten *S*s were randomly assigned to each of the eight combinations of situation and instructions, with the constraint that half of the *S*s in each condition were males and half were females.

Each *S* was presented with a questionnaire at the front of which were detailed instructions on the meaning of making probability estimates, including an explanation and illustration of independent and nonindependent events. The *S*s were told either to assume that events in the experimental task were independent or nonindependent.

For the chips situation, *S*s then received the following instructions:

> For the following questions assume that you are observing chips being drawn from a bag, in which there is a very very large number of chips. Each chip drawn may either be blue or white. If, in fact, the bag is what we will call the "blue bag," assume that 70% of the chips in the bag will be blue. If, in fact, the bag is what we will call the "white bag," assume that 70% of the chips will be white.

Instructions for the other conditions were identical, except that other hypotheses and events were substituted for the color of chips. For the military situation, the two possible hypotheses were that Russia was preparing to go to war or that Russia was preparing for peace; the two possible kinds of events were either a warlike move (mobilization of military forces in a given part of the world) or a peaceful move (demobilization of military forces). For the group situation, the two possible hypotheses were that the group was one *S* liked or that the group was one the *S* disliked; the two possible events were observing that a particular member of that group was attractive (described by a positive adjective) or unattractive (described by a negative adjective). For the individual situation, the two possible hypotheses were that the individual was one *S* liked or one *S* disliked; the two possible events were observing that a given trait of that individual was positive or negative.

---

[1] This research was partially supported by National Science Foundation Grant GS-28178. We are indebted to Ellen Reser for her assistance in coding the data and to Lawrence Becker for his help in analyzing it.

[2] Requests for reprints should be sent to Philip Brickman, Department of Psychology, Northwestern University, Evanston, Illinois 60201.

In each condition, *S*s were first asked to assume that one hypothesis was true and to predict the probability of observing a particular event on the next trial. The *S*s were asked to assess this probability for each possible configuration of prior evidence in a two-item situation (either 0, 1, or 2 prior confirming events), and for each of five possible configurations of prior evidence in an eight-item situation (either 0, 2, 4, 6, or 8 prior confirming events). The *S*s were then asked to make these same predictions for each of the other three possible combinations of hypothesis and event.

Results.—An analysis of variance was carried out on the conditional probability estimates using situations and instructions as between-*S* variables and amount of prior confirming evidence and nature of event (confirming or disconfirming of the hypothesis being considered) as within-*S* variables. Since results for the two-prior-events and eight-prior-events situations were similar, only those for the eight-prior-events situation will be reported (see Table 1). Independence versus nonindependence instructions had no effects; thus, the data in Table 1 are presented collapsed across instruction conditions.

Overall, the number of prior confirming events had strong opposite effects on estimates for subsequent confirming versus disconfirming observations; this interaction was highly significant with an $F(4, 288) = 76.43$, $p < .001$. The probability for a subsequent confirming event was elevated when few confirming events had occurred previously and depressed when many confirming events had occurred previously. The probability for a subsequent disconfirming event was depressed when few confirming events had occurred previously and was elevated when many confirming events had occurred previously.

There was also a significant triple interaction between inference situation, prior confirming events, and nature of subsequent event, $F(12, 288) = 5.76$, $p < .01$. The negative recency effect was stronger in the three real-life simulation situations than in the poker-chips situation. Simple-effects analyses indicated a significant overall trend in all cases except for disconfirming events in the poker-chip situation. A test for linear trend was significant in all conditions including this poker-chip case.

TABLE 1

Conditional Probability of Confirming and Disconfirming Events as a Function of Situation and Prior Evidence

Situation/ event	Prior confirming events				
	0/8	2/8	4/8	6/8	8/8
Chips					
Confirming	77.6	72.5	70.5	70.7	62.7
Disconfirming	33.3	32.5	33.1	38.6	38.9
Military					
Confirming	85.0	77.5	70.9	65.1	59.6
Disconfirming	21.0	27.6	30.7	34.5	41.1
Groups					
Confirming	86.7	76.8	67.5	56.4	39.2
Disconfirming	18.8	30.6	36.0	39.8	59.1
Individual					
Confirming	84.2	73.2	67.9	60.2	47.8
Disconfirming	20.8	30.6	33.0	34.9	52.2

Discussion.—Given their instructions, in which they were actually told a P(D/H) value, *S*s should logically have simply restated that P(D/H) value each time such an estimate was subsequently requested. Yet they did not; instead, as predicted, *S*s varied their estimates as a function both of the inference situation and the nature of prior evidence. In circumstances which have previously been found to yield most conservatism, *S*s were most likely to underestimate the conditional probability for another confirming event as the amount of prior confirming evidence increased.

Why did *S*s behave like this? Were they confused? Simply because *E* asked them to estimate P(D/H), *S*s may have felt that *E* expected them to vary their answers from the given value. However, since *S*s are generally not likely to go along with an experimental demand at the cost of making themselves look bad or incompetent (Weber & Cook, 1972), it would seem that they must still have felt that they were doing the right thing.

The present data seem to indicate that misperception of the basic parameters of a Bayesian model does take place, although this does not of course rule out the fact that misaggregation of the parameters may also occur (Edwards, 1968). The present results may also contribute to the explanation of why conservatism is typically found to be least on the first trial: on that trial, with no prior events, P(D/H) should be least likely to deviate from its initial optimal value. In a more indirect fashion, this study may also help to explain the inertia effect or the primacy effect in probabilistic inference. Geller and Pitz (1968) discovered that *S*s were likely to reduce their confidence in a hypothesis when a disconfirming event unexpectedly occurred, but not to reduce this confidence (i.e., to display the inertia effect) when they had predicted the observed disconfirming event. It follows that the enhanced tendency to predict disconfirming events as initial evidence accumulates in favor of one hypothesis could serve as a means by which *S*s managed to discount such disconfirming events when they occurred.

In general, the reasoning and the trend of the results of this experiment seem compatible with the argument by Kahneman and Tversky (1971) that *S*s distort their probability estimates in order to make samples be as "representative" as possible of their population. The population proportion was always given as 70%. To make the sample representative, *S*s should overpredict a confirming event when fewer than 70% confirming events have been observed, and underpredict a confirming event when more than 70% confirming events have been observed. In general, this is what they do. But contrary to strict representativeness, they seem most likely to assign a 70% probability to a subsequent confirming event when prior evidence has split 50–50 (in the 4/8 case), rather than when prior evidence has been 75% confirming events (in the 6/8 case).

Finally, the absence of an effect for independence versus nonindependence instructions is puzzling, and should not be taken as conclusive. The variations in P(D/H) suggest that *S*s in general treated events as nonindependent, regardless of instructions. Since most nonlaboratory inference situations involve multiply correlated events, it should not be too surprising to find that people were unable to abandon this assumption even when instructed to do so. In turn this bias may contribute toward their tendency to regard events as less diagnostic than does an ideal operator, who considers the events to be conditionally independent.

REFERENCES

Alker, H. A., & Hermann, M. G. Are Bayesian decisions artifically intelligent? The effect of task and personality on conservatism in processing information. *Journal of Personality and Social Psychology*, 1971, **19**, 31–41.

Edwards, W. Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment*. New York: Wiley, 1968.

Geller, E. S., & Pitz, G. F. Confidence and decision speed in the revision of opinion. *Organizational Behavior and Human Performance*, 1968, **3**, 190–201.

Jones, M. R. From probability learning to sequential processing: A critical review. *Psychological Bulletin*, 1971, **76**, 153–185.

Kahneman, D., & Tversky, A. Subjective probability: A judgment of representativeness. *Oregon Research Institute Bulletin*, 1971, **11**, No. 2.

Weber, S. J., & Cook, T. D. Subject effects in laboratory research: An examination of subject roles, demand characteristics, and valid inference. *Psychological Bulletin*, 1972, **77**, 273–295.

(Received August 6, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 237–239

VISUAL RECOGNITION OF SIMILARITY AND IDENTITY[1]

PETER L. DERKS[2]

College of William and Mary

Two groups of six *S*s compared pairs of line patterns for a match between any pair of lines in the two patterns (similarity) or for a match between all pairs of lines (identity). The similarity match took longer than the identity match, and the difference in latency increased as the number of lines in the pattern increased. Thus, R. W. Sekular and M. Abrams' 1968 report of serial processing for similarity and unitary processing for identity was replicated and extended.

Sekular and Abrams (1968) compared decision latencies between *S*s who judged 4 × 4 matrices for similarity (any cells blackened in common) or identity (all cells blackened in common). One, two, or four cells were blackened in each matrix. The latency for similarity judgments increased with an increase in the number of blackened cells, while identity response time remained fairly constant over the range tested. This performance indicated that similarity judgments were made by sampling one element at a time (serial processing), while identity judgments could be based on groups of elements (parallel processing). The present study extended this phenomenon to more letter-like stimuli, excluding actual letters, and increased the number of elements in the stimulus pairs.

Method.—The lines that composed the arrays were presented on two adjacent panels (Alpha-Numeric Electroluminescent) from Massey-Dickinson. Each panel contained 14 lines (.47 cm. wide and from 2.19 cm. to 3.13 cm. long) plus a dot (.63 cm. in diameter) in the lower right-hand corner. These 15 elements could be independently illuminated and were arranged in a configuration that could produce alpha-numeric arrays. The brightness of the elements was adjusted for each *S* so that the light blue of the illuminated elements was judged to "stand out clearly" against the dark background of the panel and the unlit elements. Positions of the unlit elements were also faintly visible. The *S*'s eyes were typically about 60 cm. from the panels.

The two patterns were displayed simultaneously and always contained an equal number of elements. The number of elements in a single pattern on any experimental trial varied from one to seven. The number of elements that matched between any two patterns was varied from none to all elements the same. For larger numbers of elements, alternate numbers of matching elements were tested between all and none. For example, for seven elements, zero, two, four, six, or seven elements were matched.

Figure 1 gives examples of pairs of stimuli with six elements to illustrate *S*'s task. The relevant conditions to test for sequential and parallel processing are the "NO" responses for similarity (upper right in Fig. 1) and the "YES" responses for identity (lower left in Fig. 1). In both cases, all elements must be examined before an accurate decision can be made. Examples in which a single element leads

---

[1] Thanks to Chris Cheney for her assistance in the data analysis and Lynn S. Schulz for her suggestions.

[2] Requests for reprints should be sent to Peter L. Derks, Department of Psychology, College of William and Mary, Williamsburg, Virginia 23185.

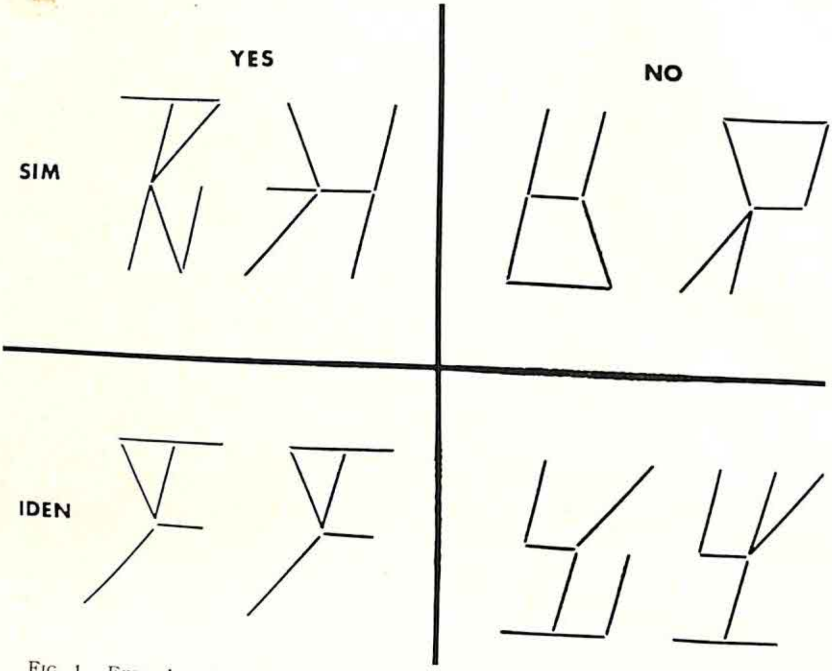


FIG. 1. Examples of possible pairs of patterns and the correct responses for each condition. (For similarity (SIM) the YES response is determined by the matching vertical line in the lower-right corner of the two patterns. Since no elements match in the other six-element pair illustrating the similar condition, the appropriate response is NO. For identity (IDEN), YES occurs when all six elements match between pattern pairs but the NO results because the lower-right vertical line in the first pattern of that square is missing from the second pattern and replaced by an upper-central vertical line.)

to similar–YES and identity–NO decisions are also included in Fig. 1.

Two groups of six *S*s were paid to judge either the similarity or identity of pairs of patterns. They received feedback on all trials and were instructed to be accurate and fast, in that order. The Similarity group was instructed to respond YES when the pair of patterns had any elements in common and NO if none of the elements matched. The Identity group was told to say YES if all the elements in a pair matched and NO if any elements differed. The duration between *S*'s activation of the patterns and the onset of the verbal response, which activated a throat microphone, was measured in milliseconds by a Hunter Klockounter. Activating the throat microphone also terminated the display.

Each *S* received four sessions at least 1 day apart that included three relevant tests at each number of elements. Nineteen stimuli that required the opposite response balanced the 21 relevant presentations in each session. Incorrect responses were not included in the latency analysis. When a stimulus pair elicited an error, the condition represented by that pair was retested later in the session with another randomly selected pattern.

There is evidence that YES and NO responses have different latencies, a result that would cause difficulty in comparing the relevant conditions (Bindra, Donderi, & Nishisato, 1968). In the present experiment, however, when *S*s were given preliminary training with the individual elements, receiving a randomization of 15 presentations with each element paired once for YES and 15 presentations with each element presented once on each panel and not paired for NO, the average reaction time for YES was .544 sec. and for NO was .524 sec. The differences between the latencies is small and not significant. There was also no evidence that any element or pair of elements was exceptionally difficult to discriminate or compare. The *S*s were matched and assigned to the different judgment conditions on the basis of their performance on the single-element task.

Results and discussion.—Figure 2 shows the average reaction time (RT) for the judgment conditions and the patterns of the present experiment. There are 18 conditions that have the same number of elements and matching elements for the Sekular and Abrams' (1968) experiment and the present one. The RTs for these conditions in the present experiment were all within the range of the RTs for their various replications of the same conditions. Thus, the agreement between comparable conditions in the two experiments was quite good.

Similarity judgments with no elements matching took longer than identity judgments with all elements on equivalent patterns of elements, $F(1, 10) = 60.61$, $p < .001$. As the number of elements increased, the differences between latencies for the two groups increased, $F(6, 60) = 8.62$, $p < .001$. This interaction indicates that the kind of judgment influenced the rate of element processing. The increase as a function of number of relevant elements was also significant, $F(6, 60)$

$= 20.24$, $p < .001$, due primarily to the increase in latency for similarity judgments. The Scheffé allowance calculated to assess differences between latencies for various numbers of relevant elements was .266 sec. for $p = .05$ and .303 sec. for $p = .01$ (Myers, 1966). The increase in latency for identity judgments over the range from one to seven elements was .216 sec. The linear function for the similar judgment latency fit by least squares was: RT (in seconds) $= .247$ sec. $+ .346\,N$ (number of elements) sec. Thus, each additional element caused a significant increase in RT for similarity. Similarity judgment resulted in relatively slow response times and indicated sequential processing of elements. The search for identity was faster and, for five, six, and seven elements in the patterns, apparently accomplished in parallel since there was no increase in RT with an increased number of elements. Nevertheless, the .054 sec. per element increase in identity judgments for from one to five elements is within the range of consistent sequential effects reported in other visual comparison studies (for example, Posner & Mitchell, 1967).

For similarity judgments, the sequential search interpretation is further supported here and in the Sekular and Abrams (1968) study by decreased latencies as a function of an increase in matching elements. The more matching elements there were in a pair of patterns, the sooner *S* found one and responded. When all elements matched, i.e., identical stimuli only judged for similarity, the YES response had about the same latency as the YES identity judgments.

For identity judgments, a single nonmatching element increased the NO RT relative to the YES RT by about the same amount as a single nonmatching element slowed YES similarity judgments for the same pattern types. Additional nonmatching elements, however, did make NO identity responses as fast as or even slightly faster than YES.

The parallel between similarity and identity RTs for patterns that were identical or differed by only one element supports the concept of a dual processing system proposed by Bamber (1969). He has suggested that there is a fast identity reporter and a slower serial processor. The identity processor may be focused on the stimulus as a whole while the serial processor subsequently examines the elements of the stimulus. The present data suggest that the serial processor's response to single elements in relatively complex arrays requires about the same time as the response to the identity of the arrays as wholes. Thus, these results also suggest that *S*s do not "adopt distinctive modes of pattern recognition [Sekular & Abrams, 1968, p. 238]." Instead, *S*s may process the patterns in the same way regardless of instructions, beginning with the determination of whole pattern identity and continuing with element-by-element analysis. Instructions simply determine which phase of the process *S*s use for the response.

Whatever the nature of processing, the difference between identity and similarity judgments was not a function of an adjusted error criterion. For identity responses, with perfect matching and YES as the relevant condition, the miss rate was 5.4%. For similarity responses, with no matching elements and NO as the relevant condition, the false-alarm rate was 6.2%, slightly higher and not significantly different. Variability in the number of nonmatching elements probably caused the higher false-alarm rate for identity, 14.1%, and a higher miss rate for similarity, 10.2%. Most of these errors occurred in patterns with several elements in the pattern and only one pair of elements determining nonidentity or similarity. Although this single element determination occurred more frequently in the identity condition, the error rate for identity was lower, though not significantly, and the RT was faster. Identity judgments could be accomplished with something like a "single look" at the whole stimulus, while similarity judgments required further "looking back and forth" to analyze the stimulus into elements.

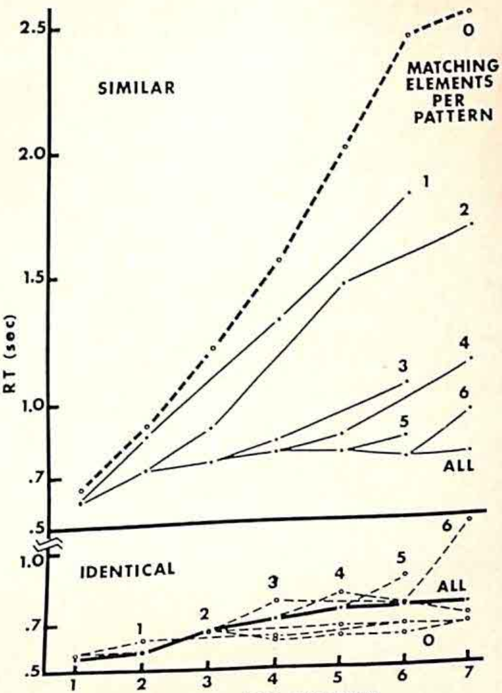


FIG. 2. Reaction time (RT) for Similarity and Identity as a function of total elements per pattern and number of matching elements.

REFERENCES

BAMBER, D. Reaction times and error rates for "same"–"different" judgments of multidimensional stimuli. *Perception & Psychophysics*, 1969, **6**, 169–174.

BINDRA, D., DONDERI, D. C., & NISHISATO, S. Decision latencies of "same" and "different" judgments. *Perception & Psychophysics*, 1968, **3**, 121–130.

MYERS, J. L. *Fundamentals of experimental design.* Boston: Allyn & Bacon, 1966.

POSNER, M. I., & MITCHELL, R. F. Chronometric analysis of classification. *Psychological Review*, 1967, **74**, 392–409.

SEKULAR, R. W., & ABRAMS, M. Visual sameness: A choice time analysis of pattern recognition processes. *Journal of Experimental Psychology*, 1968, **77**, 232–238.

(Received August 13, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 240–241

TRANSFER FROM FREE-RECALL TO PAIRED-ASSOCIATE LEARNING

A. KEITH BARTON[1] AND ROBERT K. YOUNG

University of Texas, Austin

Each of 60 *S*s first learned a free-recall list composed of four categories of six items each and then learned a paired-associate (PA) list. The *S*s in the Experimental groups learned a PA list composed of items taken from the free-recall list while *S*s of the Control groups learned two unrelated lists. The SC (Same Category) group learned a PA list such that both members of each pair belonged to the same category. The DC (Different Category) group learned a PA list such that both members of each pair belonged to two different categories. Negative transfer was found with greater negative transfer occurring in the DC group than for the SC group. Alternate explanations were discussed.

Postman (1971) recently found that under a variety of conditions negative transfer occurs from free-recall to paired-associate (PA) learning. The negative transfer observed here was consistent with Postman's position that associative interference should occur if second-list (PA) learning requires disruption of first-list subjective groupings and the establishment of new associations.

The present article represents an extension of Postman's (1971) research on transfer from free-recall to PA learning. In this study the arrangement of the PA pairs was manipulated such that the subjective groupings formed in free recall were either compatible or incompatible with subsequent PA learning. This was done by having *S* learn by free recall a list composed of several categories of words. Subsequently, *S* learned a list in which each pair was composed of items either taken from the same category or taken from two different categories. Since category clustering is common (e.g., Bousfield, 1953) S–R pairs constructed from two different categories should have both members in different subjective groupings while S–R pairs constructed from the same category should have both members in the same subjective grouping.

Method.—Each of 60 *S*s learned a free-recall list composed of four categories of six items each for five trials and then learned a PA list composed of 12 pairs for eight trials. For the Experimental (E) groups the two lists were composed of the same items and for the control (C) groups the lists were composed of completely different items. The PA list was constructed so that both members of each pair belonged to the same category (SC group) or belonged to two different categories (DC group). The SC group was further divided on the basis of blocked (SCB group) or random (SCR) presentation of the category pairs during study trials in PA learning. Thus the basic design is a 2 × 2 factorial with an additional variable, blocked versus random presentation, nested within the category variable ($n = 10$ per cell).

Five alternating study and test trials were given during free recall. A 3-sec. rate of presentation was used on study trials and a 30-sec. recall interval was used during test trials. A 6-sec. intertrial interval (ITI) was used. The responses were written in five-page booklets with one page being used per trial. Three orders of presentation were used to reduce the possibility of serial learning.

A PA list of 12 pairs was formed from the free-recall lists learned by Group E such that both C and E groups transferred to a common PA list. Two such lists were formed—one for the DC group and one for the SCB and SCR groups. The PA lists were presented for eight alternating study and test trials. Order of presentation of the stimulus terms on the test trials was random and the same for the DC, SCB, and SCR groups with the differences between the groups being apparent only on the study trials. Presentation rate was 3 sec. for both study and test trials with an ITI of 6 sec. On test trials, as the stimulus appeared, *S*s were instructed to write the appropriate response in an eight-page booklet. Three presentation orders were also used during PA learning.

The verbal materials employed in the present experiment were taken from Battig and Montague (1969). Six of the seven most frequent responses were taken from each of eight categories. Items from four of the categories were used in the free-recall list of the Control conditions and the items from the remaining four categories were used in the free-recall list of the Experimental conditions and in the PA lists. In the PA lists of the SC groups, both members of a pair were exemplars of the same category, e.g., ORANGE-PEAR, LAMP-COUCH, ENGLAND-GERMANY, etc. In the PA lists of the DC groups, the members of each pair belonged to different categories, e.g., LION-CHERRY, BANANA-ITALY, COUCH-HORSE. The pairs were arranged such that three words from each of the four categories served as stimuli and the remaining three words per category served as responses.

All words were presented by a Kodak 650 Carousel slide projector with attached timer. The *S*s were run in groups of five and were naive to verbal

---

[1] Requests for reprints should be sent so A. Keith Barton, Department of Psychology, University of Texas, Austin, Texas 78712.

TABLE 1

MEAN CORRECT RESPONSES GIVEN DURING PA LEARNING

terion	Cond.	Group					
		DC		SCR		SCB	
		M	*SD*	*M*	*SD*	*M*	*SD*
Trial 1	E	2.10	1.73	5.60	2.50	4.30	2.16
	C	5.80	2.74	6.00	1.49	6.70	1.49
All trials	E	58.80	19.32	70.10	11.75	64.40	13.56
	C	77.80	10.80	75.90	15.91	71.60	12.72

Note.—Abbreviations: E = experimental, C = control, DC = different category, SCR = same category random, SCB = same category blocked.

learning experiments. No *S*s were dropped from the experiment.

Results and discussion.—Number of correct responses given during free recall was analyzed and no significant differences were observed. In addition, a modified ratio of repetition score (Bower, Lesgold, & Tieman, 1969) was analyzed to assess clustering performance. There was considerable clustering in both the Experimental and Control lists with MRR scores of .79 and .82, respectively, but again no significant differences were found. Thus the hypothesis of equivalent groups cannot be rejected.

The mean numbers of correct responses given during PA learning are presented in Table 1. In the analysis of these data, negative transfer was found with $F\ (1, 54) = 9.49$, $p < .01$. The absence of a significant interaction, $F\ (1, 54) = 1.46$, $p < .20$, suggests that the amount of negative transfer observed in the SC and DC groups is approximately the same. However, a Newman-Keuls test found significant negative transfer only in the DC groups. A significant trials effect was also found with $F\ (7, 378) = 91.70$, $p < .001$. In addition the E versus C $\times$ Trials interaction was significant with $F\ (7, 378) = 3.31$, $p < .01$, such that the amount of negative transfer decreased over trials. Positive transfer was never found on any trial and the significant interaction suggests that a ceiling effect was operating. No other *F*s were significant in the analysis.

Number correct during the first PA trial was also analyzed. These data are also presented in Table 1. Again, negative transfer was obtained with $F\ (1, 54) = 16.30$, $p < .001$. In addition, groups and the Groups $\times$ C versus E interaction were significant with $F\text{s}\ (2, 54) = 4.56$ and 3.20, respectively, $p < .05$. The significance of the interaction suggests that greater amounts of negative transfer occurred early in PA learning in the DC groups than in the SC groups. Again, the Newman-Keuls test found significant negative transfer only in the DC groups.

Negative transfer was found in the DC groups, replicating Postman (1971), with smaller and insignificant amounts of transfer being found in the SC groups. This latter finding provides marginal support for a suggestion by Postman that negative transfer would be found in the SC groups. No positive transfer was found in the SC groups as might be expected from the learning of S–R pairs constructed from items belonging to the same subjective grouping. One reason, as suggested by Postman, for finding evidence only for negative transfer in the SC groups is that during free recall *S* has alternate pathways or associations (within a category) while during PA learning *S* must learn to use one association exclusive of all others. That is, positive transfer was not found in the SC groups because a stimulus could identify the category of the response but not the response itself. From the previous free recall the five other items of the category would be readily available, with the task of *S* being to learn which of the five was the appropriate response. Under such conditions facilitation would be expected in the SC groups only when two-member category lists were learned so that in PA learning use of one member of the category as the stimulus would uniquely identify the appropriate response. Even here it is doubtful if positive transfer would be obtained. The reason for this is that the interitem associations (cf. Deese, 1959) which exist between members of a category would facilitate the learning of the control group list so much that it would be difficult for the experimental group to learn it any faster.

Consistent although insignificant amounts of negative transfer were found in the PA learning of the SC groups suggesting that interference did occur. In addition the data also support the conclusion that greater amounts of negative transfer occurred in the DC than in the SC groups. The reason for this may be due to factors already discussed: Selection of the appropriate response in the SC groups requires that the strength of an already existing interitem association be increased sufficiently to be discriminable from competing interitem associations. Selection of an appropriate response in the DC groups, on the other hand, would require establishment of an entirely new association.

REFERENCES

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 1969, **80** (3, Pt. 2).

Bousfield, W. A. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 1953, **49**, 229–240.

Bower, G. H., Lesgold, A. M., & Tieman, D. Grouping operations in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 481–493.

Deese, J. Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, 1959, **5**, 305–312.

Postman, L. Organization and interference. *Psychological Review*, 1971, **78**, 290–302.

(Received August 20, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 1, 242–244

STIMULUS RECOGNITION AND ASSOCIATIVE CODING

WILLARD N. RUNQUIST[1] AND ANNABEL EVANS
University of Alberta

The *S*s learned CCC-noun pairs by the study-test method, where test trials demanded *S* select the relevant stimuli from a list and recall the response. For three groups the initial letter of the response term was contained in its trigram. Two of these groups were told about this relation with one group having all code letters in the same position, while in the second group their position varied from item to item. The third group had a consistent position but were not told about the compatible code. A fourth group learned a list in which the items within a pair were unrelated. Recognition performance was superior in the consistent position-instructed group, but did not differ among the others. Recall given recognition was equivalent for the instructed groups, and better than that in the noninstructed groups. The noninstructed group with compatible response terms did recall better than the unrelated group.

The purpose of this experiment was to investigate the relationship between stimulus recognition and various learning conditions which were designed to affect both stimulus encoding and associative learning in a paired-associate task. These conditions involved using pairs in which the initial letter of each response word was contained in its trigram stimulus. By informing *S*s that this relation exists, it should be possible to manipulate the stability of stimulus codes by varying the consistency with which this "code letter" appears in the same position in the various stimuli.

The experiment should bear upon two theoretical problems discussed by Martin (1968): the effect of variability of stimulus codes upon recognition performance, and the independence of coding and associative processes in learning.

Method.—Each *S* was given eight study-test trials on a six-item paired-associate list in which the stimulus terms were CCC trigrams with no letters duplicated and the response terms were one-syllable common nouns. Four groups of 20 *S*s from an introductory psychology course were run. Assignment to three of the groups, code consistent (CC), code inconsistent (CI), and neutral (N) was done according to a prearranged scheme of blocks containing each condition in a random order. The fourth group, code consistent-not instructed (CC-N), was run after the completion of the remaining three groups. Each group learned a list composed of the same stimuli. For Groups CC and CC-N, the first letter of the response term was the same as the middle letter of the CCC. For Group CI, the first letter of the response term was the same as the first letter of the CCC for two pairs, the same as the middle letter for two pairs, and the same as the third letter for two pairs. For Group N, the first letter in the response term did not correspond to any letter in its paired stimulus. In addition to list differences, the groups were also differentiated by instructions. Groups N and CC-N were given standard paired-associate instructions. Groups CC and CI were told in addition that it was possible to relate the stimulus and response terms by identical letters and were shown an example pair.

On study trials, *S*s were presented each pair for 2 sec. For test trials, they were given a sheet containing 18 CCC trigrams each followed by a blank space. They were asked to circle each stimulus term from the list and write the appropriate response term next to it. All six stimuli appeared on these sheets on each test trial along with 12 new "distractor" items. The nature of the distractors was not systematically manipulated, but was the same for all groups. Because 18 of the 20 possible consonants (excluding Y) were used in constructing stimuli, all distractors contained letters common to the stimuli. Six of the items on each trial had no more than 1 consonant from any one trigram, and always in different positions. A few items contained the 2 consonants not in the list. Three items on each trial contained the code letter from three of the items in the CC group. In addition, three more items contained one first-, one second-, and one third-position code letter from the CI groups. The remaining consonants were all from different positions and different items. Order of appearance was random and varied from trial to trial, but was the same for all *S*s. Eight separate presentation orders were used on study trials. Unlimited time was allowed for *S* to complete the test trial.

Prior to learning the test list, each *S* was given one study trial on two pairs (TOM-FLIES, BOB-CLIMBS) followed by a recognition recall test on seven names.

Results.—A plot of the trial-by-trial data showed no theoretically interesting trial interactions; thus, overall percentage of correct recognitions, percentage of correct recalls, and percentage of recall given correct recognition are shown in Table 1.

Recognition performance appears to be superior for the CC group. The other groups, however, differed little from one another. Analysis of variance produced a significant $F\ (3, 76) = 3.23$, $p < .05$, and Duncan's multiple-range test verified the above pattern.

---

[1] Requests for reprints should be sent to W. N. Runquist, Department of Psychology, Univeristy of Alberta, Edmonton, Alberta, Canada.

TABLE 1

Percentage of Correct Recognitions, Correct Recall, and Recall Given Recognition

Group	Recognition	Recall	Recall/ Recognition
N	79.5	52.8	65.2
CC	87.3	85.8	97.5
CC-N	78.9	71.9	84.2
CI	78.1	75.3	95.1
Est s_m	2.40	2.86	3.03

Overall recall also differed for the various groups, $F(3, 76) = 23.21$, $p < .01$. The multiple-range test revealed that Group CC performed better than all other groups, while Group N performed more poorly. Groups CI and CC-N did not differ on this measure.

When recall was conditionalized upon recognition, a somewhat different pattern emerged. The differences among groups were also reliable, $F(3, 76) = 23.52$, $p < .01$, and the multiple-range test indicated that Groups CC and CI exhibited superior performance to Group CC-N and to Group N, while the latter group recalled fewer recognized items than any other group.

Also of some interest are the recognition errors (false alarms), a summary of which is presented in Table 2. Analysis of variance of mean total errors (column 1) showed that the frequency of false recognitions differed among the groups, $F(3, 76) = 7.75$, $p < .01$. The multiple-range test revealed that Group CI made more errors than all other groups which did not differ from one another.

The classification of errors in Table 2 was made in two ways. Columns 2, 3, and 4 indicate the nature of the false recognitions. Column 2 shows the frequency of recognition errors which contain the first letter of one of the response terms, even though that letter is in a different position in the trigram. Column 3 presents the frequency of errors in which other letters from the stimulus appear in the same position, and the fourth column indicates the frequency of false recognitions in which all letters are in a different position or not in the list at all. The results show that despite invariance in the number of errors for Groups CC, CC-N, and N, there are differences in the kinds of errors made. Almost all of the errors in Group CC are those containing the code letter, while the frequency of "unrelated" errors is higher in CC-N and N. The differentially high false-alarm rate in Group CI is exclusively a result of an increase in code letter errors.

The remaining columns deal with errors which were accompanied by an attempted recall. The total number of these errors is given in column 5. Appropriate responses (column 6) are those which are related to the code letter. That is, they represent the frequency of times *S* falsely recognized an item containing the first letter of a response and gave that response. Inappropriate errors (column 7) represent all other false recognitions accompanied by an attempted recall. Again despite lack of differences in total misrecognitions, the pattern differs. Group CC invariably attempted recall and gave an appropriate response. Groups CC-N and N had a large proportion of omissions, although the percentage in this category was much higher in Group N. The large number of code errors in Group CI was accompanied by a large number of appropriate responses.

Discussion.—The most interesting comparison in this experiment involves Group CC and Group CI. Group CC in which the response terms were consistently related to a single letter position in each stimulus produced both better stimulus recognition and overall recall than Group CI in which the response terms were related to different letter positions in each stimulus. When recall is conditional upon recognition, however, the two groups did not differ.

If the development of a stable stimulus encoding is necessary for recognition, one might interpret the facilitation of recognition in the CC group as resulting from greater encoding stability with the parenthetical note that it is the associative code which provides this stability. Given that stable recognition is established, however, the formal relation between stimulus and response terms facilitates recall equally. Thus, in a somewhat indirect fashion, these findings are consistent with the proposition that coding of the stimuli (as indexed by stimulus recognition) and the associative process are essentially independent.

Further considerations, however, suggest that this conclusion be held with some reservation. It could easily be argued that successful performance in the recognition task in this experiment placed demands on *S* which are not necessary for successful associative recall. In both CC and CI conditions, a single letter is logically sufficient as a recall cue. Thus, *S*s who utilize this as a functional stimulus could mediate recall successfully and yet fail to know whether the remaining two letters were indeed in

TABLE 2

Frequency of Recognition Errors

Group	(1) Total	(2) Code	(3) Other	(4) Unrelated	(5) With error	(6) Appropriate	(7) Inappropriate	(8) Omission
N	44	—	25	19	11	4	7	33
CC	40	39	0	1	39	37	2	1
CC-N	56	29	8	19	32	22	10	24
CI	107	96	3	18	89	74	15	18

the list. Thus, we may be penalizing *S*s for "doing the right thing" by considering only the recall of items they recognize completely. A more fair base line for CC and CI *S*s might be achieved by including those items which *S* falsely accepts but which contain the code letter. Table 2 shows that there were 39 such items in the CC group, 37 of which were accompanied by an appropriate response, while in the CI group, 96 recognition errors were accompanied by 74 appropriate responses. Considering these items as correctly recognized stimuli, the conditional probabilities are .973 for CC and .864 for the CI group. Moreover, the result is the same whether the sample includes items in which the code letter was in the correct position, or whether it was in some other position.

Just why this result should obtain is not entirely clear. Both CC and CI *S*s were instructed concerning letter–response relations. Thus, if *S* selected a letter based on the response, then recognized the letter, even though it was combined with inappropriate other letters, he certainly should be able to generate the response. Furthermore, there is little reason to expect intralist code position variability to affect the availability of responses (in the response learning sense). It may very well be the case that the variability in the position of a potential code letter hindered the associative phase in an indirect manner by leading *S*s to abandon the code letter in favor of some other letter or attribute which was less effective as a retrieval cue. Just why instructed *S*s should be so perverse is not clear. The only evidence supporting this latter possibility is that the CI group also made a large number of false recognitions of items not containing a code letter and occasionally recalled a response which was compatible with one of the letters in those items. It is also possible that these errors indicate the use of a different recognition criterion (in the signal detection theory sense) on the part of *S*.

The findings with respect to Group CC-N may also be commented on briefly. The recognition performance of this group was quantitatively similar to Group N, indicating that the instructions are necessary if the formal relation between S and R is to facilitate stimulus recognition. Given recognition, however, some facilitation in association is clearly apparent, with this probably depending upon whether *S*s did in fact discover the code-response relation. The recognition error data clearly indicate this to be the case, as 22 of the 29 recognition errors containing the code letter were accompanied by appropriate responses. These results also indicate that while both the number of correct recognitions and false recognitions did not differ greatly between Groups CC-N and N, the actual recognition behavior differed considerably. When Group N made a recognition error, it was often based on a common letter. These letters infrequently mediated an appropriate recall, or any recall at all. On the other hand, many of the false recognitions in Group CC-N were appropriate recalls.

Perhaps the best lesson to be obtained from this experiment is methodological. Certainly one can gain a great deal of control over letter-selection coding with these operations. However, recognition-recall tests should probably be used with some caution, for the demands of the recognition test may sometimes prove to be irrelevant to recall, at least in the ordinary paired-associate task. Cued recall outside the laboratory is probably another matter.

REFERENCE

MARTIN, E. Stimulus meaningfulness and paired-association transfer: An encoding variability hypothesis. *Psychological Review*, 1968, **75**, 421–441.

(Received September 27, 1971)

Our new

DIGITAL MILLISECOND CLOCK COUNTER

does so much more

...and at such little *cost!*

- Bright, instantly legible 4-digit L.E.D. readout.
- Timing range: 0 to 99.99 seconds.
- Counts events to 9999.
- **Floating decimal point** enables timed intervals to always be displayed in their most accurate form.
- Time base accuracy: ±.005%
- **Automatic reset.** When in "normal" mode, clock automatically begins next reading at 0. Manual reset also provided.
- **Store** mode enables unit to retain last reading until completion of next reading.
- **Accumulator** mode allows user to automatically add consecutive timings for a cumulative reading.
- Internal fifth digit rounds off to scientific notation in all ranges (all readings ±.5 milliseconds).
- Counter automatically rejects switch bounce.
- Single switch and consecutive switch operation.
- Overrange and gating indicators.
- Many additional features.
- Low cost: $385.00 f.o.b. Arlington, Massachusetts.

Contact us today for complete details.

Ralph Gerbrands Company

8 BECK ROAD, ARLINGTON, MASSACHUSETTS 02174 (617) 648-6415

"the standard of excellence in laboratories throughout the world"

HYPOTHESIS: *Harper's Texts Will Meet Your Needs*
PROOF: *Use Them and See*

Stimulus Schedules: The t-τ Systems/W. N. SCHOENFELD AND COAUTHORS: J. BLAUSTEIN, G. D. LACHTER, C. VICKERY, AND J. M. MARTIN Incorporating work done over the past 20 years by Professor Schoenfeld and his students (many of whom are his coauthors) this book presents the first systematic framework for the study of reinforcement schedules and all other schedules of stimulus presentation to a responding organism. Uniquely organized, it is divided into chapters which treat the theoretical structure of the t-τ systems, each followed by abstracts of the experiments which were generated by the systems and which will be the basis of future research. Abstracts are annotated with all pertinent data. Graphs. Bibliographies. September, 1972. Paper. Tentative: 224 pp.; $4.95.

The Experimenter's Dilemma/JOHN JUNG Integrates articles on significant developments in research methodology with an extensive discussion of general aspects and problems. Readings focus on the effects of the subject's attitude toward the experimental task; the experimenter's hypothesis on the outcome of the projects; questions regarding the representative nature of subject populations; and issues involved in such techniques as deception. 1971. Paper. 322 pp.; $4.95.

An Introduction to Experimental Design in Psychology: A Case Approach/HOMER H. JOHNSON AND ROBERT L. SOLSO This general introduction to experimental design in psychology emphasizes actual experiments rather than the philosophy of science or statistics. Design examples from a variety of areas—social, clinical, and educational—demonstrate the application of scientific methods to all areas of psychology. The text contains summarized experiments which illustrate design principles, several "dummy" experiments to be redesigned, and eight fully reprinted experiments to be analyzed. 1971. Paper. 224 pp.; $3.95.

Foundations of Experimental Research/ROBERT PLUTCHIK Views experimentation as a decision-making process that begins with the conception of a problem and ends with the publication of a report. The major decision areas concern such issues as the definition of key concepts, the sampling of subjects as well as stimuli, measurement, scaling, and the use of appropriate mathematical models. 1968. Paper. 290 pp.; $5.50.

A Primer of Experimental Psychology/JOSEPH LYONS Here is a supplement of methodology that is useful in both introductory and experimental psychology courses. It is written largely around specific examples drawn from areas such as animal behavior, space psychology, psychotherapy research, esthetics, hypnosis, and ESP, showing how research in these fields has resulted in controversy. 1965. Paper. 322 pp.; $5.50.

Send for Harper's psychology catalog (#CT 61)

1817

HARPER & ROW, 10 East 53d Street, New York 10022

DATA CUBES A NEW SHAPE FOR LOW COST BEHAVIOR CONTROL

The 5800 Data Cube series is the first in a family of low cost, high quality instruments designed especially for introductory, undergraduate and general research work in human and animal behavioral laboratories. Use them independently or in any combination. Each Data Cube in the 5800 series is guaranteed 100% compatible and all units with electronic circuitry are solid state with built-in power supplies.

Identical size 6" x 6" x 4" cubes can be stacked like blocks, eliminating the need for expensive racks and power rod strips. Front panels are engineered for easy operation. And color coding on binding posts is consistent. With the 21 Data Cubes now available in the 5800 series, the experimenter, lecturer or researcher can control all basic human and animal behavioral experiments inexpensively and completely.

For information on individual Data Cubes, including price list and specifications, check the following list and mail to: 5800 Data Cube Series, Lafayette Instrument Company, P. O. Box 1279, Lafayette, Indiana 47902.

■ indicator panel ■ switch panel ■ selector panel ■ data recorder ■ Power supply ■ shocker ■ stop clock ■ drinkometer ■ counter ■ timer ■ photocell control ■ predetermining counter ■ auxiliary relay ■ stimulus display ■ numeric indicator ■ power control ■ flicker fusion ■ GSR ■ probability generator ■ scrambler ■ pulse former

LAFAYETTE INSTRUMENT COMPANY
P. O. Box 1279 · Lafayette, Ind. 47902
(317) 742-4998

Lafayette

Look into Iconix
T-scope systems

$2,000 buys a three-field Iconix tachistoscope system—exposure box, lamp driver, and timer. You'd pay over $3,000 elsewhere. What's more, we offer several alternates for each unit in the system. With any of these catalog items you get maximum performance at minimum cost.

Iconix equipment is flexible! If you need projection backs, we've got them. Want an adjustable polarizer set to present dichotic fields? It's a standard accessory. And all systems are open-ended: at your option you can add paper tape programming, response storage, measurement of reaction time, or automatic data recording on punched paper tape.

Only Iconix offers you an off-the-shelf logic unit to use with your basic T-scope system. It lets you store a library of experimental programs in easily wired plugboards. May we repeat: our logic units are standard catalog items—not high-priced customized equipment.

Iconix T-scope systems provide reliable solid-state electronics at a price any undergraduate lab can afford. Get our catalog. It has the whole story. Write

**ICONIX
PRODUCTS**
Siliconix incorporated
**2201 Laurelwood Road
Santa Clara, CA 95054
Hot line for info/service: (408) 243-2424**

Exposure Box. *Available with three or four fields. Each field accepts stimuli from 6" x 9" cards, slides, or 3-dimensional objects. Positioning of stimulus cards and beam splitters is externally adjustable. Viewing hood accommodates subjects wearing glasses.*

Timer. *Six preset intervals, 4 digits, with binary-coded decimal time display; thus, you can read reaction times directly. Also available is a 10-interval preset, 5-digit timer. Both have quartz-crystal timebases for ±0.005% accuracy. Timers may be controlled by the subject, experimenter or both.*

Lamp Driver. *3- or 4-channel drivers control low-voltage, hot-cathode lamps. A pre-pulse ensures lamp firing every time with low-cost standard fluorescents. Rise and fall times are 0.2 milliseconds with no after-glow.*

Journal of
Experimental Psychology

VOL. 95, NO. 2 OCTOBER 1972

EFFECTS OF DURATION OF MASKING STIMULUS AND DARK INTERVAL ON THE DETECTION OF A TEST DISK[1]

JOHN HOGBEN[2] AND VINCENT DI LOLLO

University of Western Australia

Detection thresholds for a test disk of light presented tachistoscopically were studied in one of four temporal relationships with a masking flash of variable duration. The conditions were: (*a*) simultaneous onset of test and mask, (*b*) simultaneous offset of test and mask, (*c*) offset of test coincident with the onset of mask, (*d*) onset of test coincident with the offset of mask. In each case, duration of mask was systematically varied. The four conditions were replicated with the mask substituted by a dark interval which interrupted an adapting field of light for a variable period. Masking increased with increments in the duration of the mask and was most severe under conditions where test and mask onset were simultaneous or where a dark interval preceded the simultaneous onset of test and mask. A dark interval within the adapting field always raised the detection threshold of the test disk. The results are consonant with the view that masking is due to "on-responses" in the visual system.

Visual masking of a test disk by a flash of light is obtained when the test stimulus is followed by the masking stimulus at an interval no greater than 50–100 msec. (Crawford, 1947). As the test stimulus is moved forward in time closer to the onset of the mask, its detection threshold rises steeply, reaching a peak at a point where the test stimulus onset coincides with, or slightly follows the onset of the mask. Thereafter, as long as the mask is still present, threshold diminishes to some stable value. It is frequently found to rise again when test stimulus onset just precedes or coincides with the offset of the mask, and then falls rapidly following the offset of the masking stimulus. This series of changes, referred to in the present paper as the "Crawford masking function," describes the type of masking situation known as "masking by light" (Kahneman, 1968) where little or no interaction occurs between the contours of the test stimulus and those of the mask.

Although increments in detection threshold have typically been obtained by presenting a flash of light in close temporal contiguity with the test stimulus, threshold increments have also been reported when the test stimulus was presented during an interruption of an otherwise steady light-adapting field. Hogben (1971) found that

---

[1] This research was supported by Grant ARG-17-226 from the Australian Research Grants Committee to John Ross and Vincent Di Lollo. We are grateful to William R. Uttal and John Ross for helpful discussions. In particular, we wish to thank William R. Uttal for his help in the calibration of the equipment.

[2] Requests for reprints should be sent to John Hogben, Department of Psychology, University of Western Australia, Nedlands, Western Australia, 6009.

© 1972 by the American Psychological Association, Inc.

detection threshold increased as a function of the duration of the dark interval, irrespective of whether the test sitmulus occurred at the beginning or the end of the dark interval. Ikeda and Boynton (1965) reported a similar result with a dimming of the adapting field rather than total extinction. Acuity performance is also affected by a brief dark interval: Kahneman (1966) reported a sharp decrement in acuity when an adapting field was interrupted for the duration of presentation of a Landolt C. Holzworth and Doherty (1971) obtained results consistent with the form of the Crawford masking function in a letter-identification task where the test stimulus was presented just before an interruption of the adapting field. On the basis of their results, Holzworth and Doherty proposed a theoretical distinction between masking by light and masking by darkness.

It is the purpose of the present experiment to explore the masking effects of a brief dark interval as compared to the masking effects of a brief light flash. The first portion of the experiment examined four temporal relationships between a test disk and a large, uniform masking flash presented on a dark-adapting field. In each case, the duration of the mask was systematically varied. The four conditions were replicated in the second portion of the experiment with the exception that the light and the dark cycles of the masking field were reversed.

The second portion of the experiment is notably related to a study by Boynton and Kandel (1957), who reported a decrement in the amount of masking due to the onset of a masking stimulus when the luminance of a preceding adapting field was increased. In general terms, their explanation for this finding was that the higher the adaptation level of the visual system, the less responsive it is to the masking effects of neural "on-responses." The present experiment approaches the same problem by a different method. Whereas Boynton and Kandel manipulated the sensitivity of the visual system by varying the luminance of an adapting field, the present experiment manipulated sensitivity by varying the time between the extinction of an adapting field of fixed luminance and the onset of a masking stimulus of the same luminance.

METHOD

Subjects.—The authors, three paid volunteers—two undergraduates (one female), and a high school student—served as *Ss.*

Apparatus.—A standard Scientific Prototype (Model GB) tachistoscope was interfaced with a computer, (PDP-8/S), which performed all the timing functions. Channel B of the tachistoscope, which provided the masking flash or the adapting field, depending on the condition, was 20.5 mL. in luminance and subtended 7° visual angle horizontally by 5° vertically. For use as an adapting field, it contained a tiny black fixation point at its center. When there was a dark-adapting field (Cond. a–d), Channel 2 provided a faintly glowing red fixation point. Channel 1 contained the test stimulus, which was a centrally located, 12.5-mL. white disk subtending a visual angle of 1° on a black background.

A hand-held switch and a morse key were interfaced to the computer to allow *S* to initiate his own stimulus presentation and to make his responses, respectively.

Procedure.—The eight stimulus conditions are shown diagrammatically in Fig. 1. The four conditions shown at the left are those in which the effect of mask duration was investigated: the test stimulus was presented at either the beginning or the end of the mask, and either with synchronous onset or offset (Cond. a and b) or with zero interstimulus interval (Cond. c and d). The four diagrams at the right of the figure represent conditions in which the effect of a dark interval was investigated. In addition, there were four control conditions: (a) test stimulus presented alone on a dark fixation field, (b) test stimulus superimposed on the masking field which had the same duration as the test stimulus, (c) test stimulus presented on steady adapting field, and (d) adapting field interrupted during presenta-

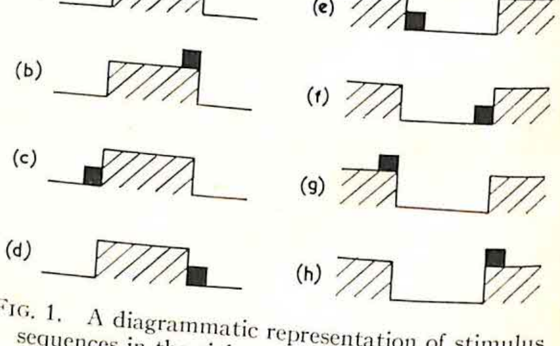


FIG. 1. A diagrammatic representation of stimulus sequences in the eight experimental conditions.

tion of test stimulus. In the execution of the two-alternative temporal forced-choice procedure (see below), the masking field in Control Cond. (b) was flashed every time, and the adapting field in Control Cond. (d) was interrupted every time, regardless of the occurrence of the test stimulus.

Under each of the eight experimental conditions, there were six masking stimulus or dark interval durations: 10, 20, 40, 80, 160, and 320 msec. In each case, the dependent variable was the duration threshold for detection of the test stimulus. The use of luminance thresholds, though generally preferable, was not possible with the equipment available. The interpretation of the results is, therefore, subject to the limitations inherent in the use of duration thresholds (see Sperling, 1965). Each *S* made four threshold estimates for each control condition and for every combination of the six mask durations and eight experimental conditions.

Thresholds were obtained by a PEST procedure (Pollack, 1968; Taylor & Creelman, 1967), which was completely controlled by the PDP-8/S computer. Briefly, the PEST procedure increased or decreased the duration of the test disk, contingent on *S*'s performance, in such a manner as to track approximately 75% correct detection. The precise PEST rules used were those employed by Pollack (1971). The response indicator was a two-alternative temporal forced choice: there were two stimulus presentations on a trial, and the test disk was illuminated on only one of these presentations. The *S* was required to indicate which one contained the disk. The lowest threshold estimate allowed by the equipment was .2 msec.

The computer used a random number generator to determine on which presentation the disk was to be illuminated and scored *S*'s responses on each trial. The response was either one or two taps on a morse key to indicate that the disk occurred on the first or the second presentation, respectively. An interval of at least 2 sec. between stimulus presentations was guaranteed in the programming of the computer. Six or seven thresholds were obtained in each 15- to 30-min. experimental session; there were 32 such sessions in all for each *S*.

RESULTS AND DISCUSSION

The four threshold measures obtained by each *S* for each combination of mask duration and masking condition were transformed logarithmically and were then averaged to yield mean threshold values for each *S*. These values, averaged over all five *S*s, are shown in Fig. 2 and 3. The mean group results shown in these figures are entirely representative of the individual results of each of the five *S*s.

Effects of masking stimulus duration.—The effect of duration of the masking stimulus was most pronounced in Cond. (a) where the onsets of the test stimulus and the mask were simultaneous. As shown in Fig. 2, threshold rises with increase in mask duration up to 80 msec., and remains unchanged with longer duration of mask. Between 10 and 80 msec. the relationship is linear with unit slope, indicating that the

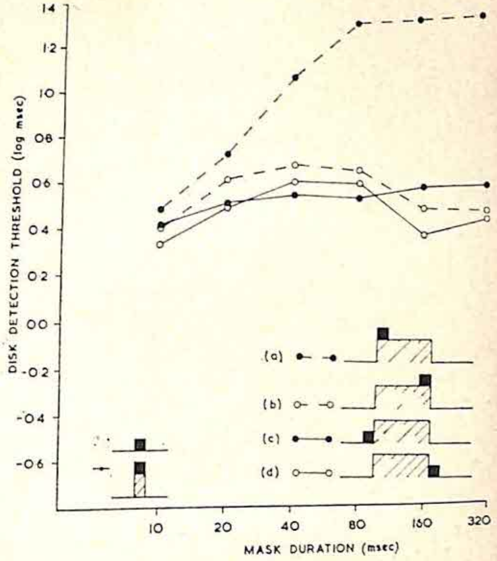


FIG. 2. Mean threshold levels for Cond. (a) to (d). (The threshold levels for both control conditions coincided at .2 msec., the lowest threshold estimate allowed by the equipment. The coincidence is thus probably artifactual.)

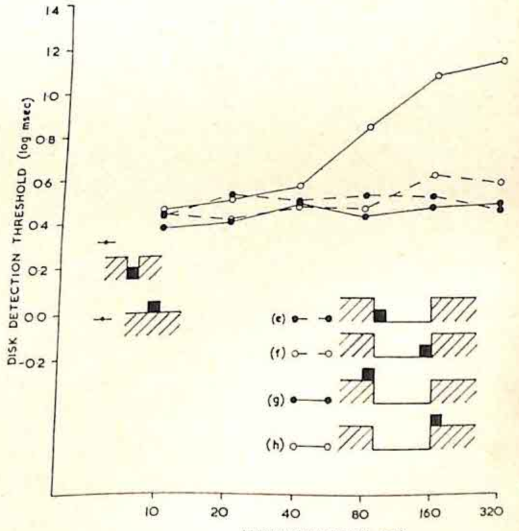


FIG. 3. Mean threshold levels for Cond. (e) to (h).

energy-contrast threshold remains unchanged as would be expected on the basis of Bloch's reciprocity law.

The curve for Cond. (a) should be compared with that for Cond. (c) where the test stimulus just preceded the mask. In both Cond. (a) and (c), the presentation of a masking flash of only 10 msec. causes an elevation in threshold of 1.0 log units as compared with control Cond. (a) where the test stimulus was presented on a dark background. Increments in the duration of the mask have only marginal effects in Cond. (c), but result in sharp threshold elevation (almost another log unit) in Cond. (a). The greater rate of threshold increment in Cond. (a) is consistent with the general form of the Crawford masking function, and with the finding that longer mask durations have little effect on the extent of backward masking but have pronounced effects on the detection threshold of stimuli following the onset of the mask (Battersby & Wagman, 1959; Sperling, 1965).

It might be argued that the discrepancy between the two sets of results is due to reduced figure-ground contrast of the test stimulus in Cond. (a) where the test disk is superimposed on the masking field, whereas in Cond. (c) the test disk is presented on the dark fixation field (e.g., Eriksen, 1966). However, this possibility is discounted by a comparison of the results for Cond. (b) and (d), where the differences in thresholds are almost negligible.

The threshold differences found between Cond. (a) and (c) may also be viewed as arising from temporal discrimination in backward masking (Sperling, 1965) precluded in Cond. (a), where test and masking fields occur simultaneously. In these terms, however, larger threshold differences between the two conditions should be expected at the shorter mask durations. In general, although the threshold differences found between Cond. (a) and (c) were expected on grounds outlined in the foregoing, the magnitude of the effect was unexpected.

This pattern of results offers clear support for the view (Baker, 1963; Boynton, 1958) that masking is largely brought about by the onset of the masking stimulus. The magnitude of the masking effect due to the onset of the mask is evidenced by a comparison of the threshold levels attained in Cond. (a) and the level found in Control Cond. (c), where the test stimulus was superimposed on a steady adapting field.

The curves for Cond. (b) and (d) are virtually identical in shape, showing maximum masking at a mask duration of 40 msec. A possible interpretation may be offered as follows: as the duration of the mask is increased from 10 to 40 msec., the magnitude of the neural responses to the mask is increased, as a manifestation of Bloch's law for time-intensity reciprocity. Above 40 msec., however, another factor must be taken into account: concomitantly with the increase in the mask duration, test stimulus onset is delayed relative to the onset of the mask, and the decline in threshold reflects the shape of the Crawford masking function.

Effects of dark interval—Presentation of the test stimulus on a dark background during an interruption of the adapting field yielded the higher of the two control thresholds shown in Fig. 3. Superimposition of the test stimulus on an uninterrupted adapting field yielded the lower threshold. Similar results were found by Kahneman (1966) using an acuity task involving the resolution of a Landolt C. Kahneman (1966) attributed his finding to an interference with the contours of the test stimulus. Threshold elevation by means of the occurrence of a dark interval is consonant with an interpretation of masking effects in terms of neural responses to the onset and offset of the mask (e.g., Baker, 1963; Ikeda & Boynton, 1965).

Conditions (e), (f), and (g) in Fig. 3 show only slightly greater thresholds than the higher control level, and increments in the duration of the dark interval from 10 to 320 msec. have very little effect on threshold. Further, for these conditions, it makes little or no difference whether the test disk is presented at the beginning or end of the dark interval.

By far the greatest effect due to the duration of the dark interval is seen in Cond. (h), where threshold increases with the dark interval up to a duration of at least 320 msec. In this condition, the onset of the test disk was simultaneous with the reinstatement of the adapting field at the termination of the dark interval; the magnitude of the effect thus obtained is comparable to that obtained in Cond. (a), where the onset of the test disk was simultaneous with the onset of the masking stimulus. It should be emphasized at this point that the adapting field of Cond. (h) and the masking stimulus of Cond. (a) were physically the same field of the tachistoscope, set to the same luminance level.

Comparison between Cond. (h) and (f) parallels that between Cond. (a) and (c). Both sets of comparisons are characterized by pronounced masking effects when the test stimulus occurs simultaneously with the onset of the masking field, and by low levels of masking when the test stimulus precedes the onset of the masking field by only a few msec. The difference in results between Cond. (h) and (f) closely parallels that between Cond. (a) and (c), suggesting the possibility that both sets of results belong to the same class of events.

In this context we hold that, contrary to the view of Holzworth and Doherty (1971), the presence of a dark interval does not of itself act as a mask. Rather, by acting upon the level of adaptation of the visual system, it enables masking effects associated with reinstatement of the light field. As shown by Keidel, Keidel, and Wigand (1961), the visual system mediates luminance discriminations over a limited range at any given time. The sensitivity range undergoes continuous adjustment according to current luminance levels, in such a way as to maximize the information gained from spatial and temporal fluctuations of luminance. Furthermore, Craik (1938, 1939) showed that performance was optimal in both an incremental threshold task and an acuity task, when the test object was superimposed on a background of the same luminance as a preceding adapting field. Performance in both tasks was poorer when the test stimulus background and the adapting field luminance differed in either direction. Thus, following the offset of the adapting field in Cond. (h), the visual system undergoes rapid dark adaptation in such a way as to maximize its discriminative capability in conditions of darkness. The longer the dark adaptation is allowed to proceed before onset of the test stimulus and mask, the less suited is the visual system to mediate discriminations involving the relatively high luminances of the test and masking stimuli. This account clearly implies the Baker–Boynton view that it is the onset of the mask that is most important in the masking of the test stimulus. In the case of Cond. (h), the onset of the mask is important because it follows a period of darkness during which the sensitivity range of the visual system has been adjusted. As a consequence, the masking stimulus onset causes a relatively large amount of neural activity. If the test stimulus is delayed with respect to the onset of the mask, the visual system has had some time to adjust to the higher luminance level, and the momentary magnitude of the neural activity due to the mask has decreased. The corresponding changes in the severity of masking produce the shape of the Crawford masking function.

The foregoing analysis of the results for Cond. (h) applies equally well to the results of Boynton and Kandel (1957), who manipulated as an independent variable the luminance of an adapting field which always terminated 280 msec. before the onset of the masking stimulus. They found that as they increased the luminance of the adapting field from zero luminance to a luminance somewhat above that of the mask, the test stimulus threshold was lowered. Further increases in adapting field luminance led to an increase in threshold. According to the account in terms of the sensitivity of the visual system, the adapting field luminance which led to the lowest thresholds was that which, after 280 msec., left the visual system within a sensitivity range appropriate to the luminance of the masking stimulus.

REFERENCES

BAKER, H. D. Initial stages of dark and light adaptation. *Journal of the Optical Society of America*, 1963, **53**, 98–103.

BATTERSBY, W. S. & WAGMAN, I. H. Neural limitations of visual excitability. I. The time course of monocular light adaptation. *Journal of the Optical Society of America*, 1959, **49**, 752–759.

BOYNTON, R. M. On-responses in the human visual system as inferred from psychophysical studies of rapid-adaptation. *A.M.A. Archives of Ophthalmology*, 1958, **60**, 800–810.

BOYNTON, R. M. & KANDEL, G. On responses in the human visual system as a function of adaptation level. *Journal of the Optical Society of America*, 1957, **42**, 275–286.

CRAIK, K. J. W. The effect of adaptation on differential brightness discrimination. *Journal of Physiology*, 1938, **92**, 406–421.

CRAIK, K. J. W. The effect of adaptation upon visual acuity. *British Journal of Psychology*, 1939, **29**, 252–266.

CRAWFORD, B. H. Visual adaptation in relation to brief conditioning stimuli. *Proceedings of the Royal Society*, 1947, **B134**, 283–302.

ERIKSEN, C. W. Temporal luminance summation effects in backward and forward masking. *Perception & Psychophysics*, 1966, **1**, 87–92.

HOGBEN, J. Interaction of forward and backward visual masking. *Perception & Psychophysics*, 1971, **9**, 487–488.

HOLZWORTH, R. J., & DOHERTY, M. E. Visual masking by light offset. *Perception* & *Psychophysics*, 1971, **10**, 327–330.

IKEDA, M. & BOYNTON, R. M. Negative flashes, positive flashes, and flicker examined by increment threshold technique. *Journal of the Optical Society of America*, 1965, **55**, 560–566.

KAHNEMAN, D. Time-intensity reciprocity under various conditions of adaptation and backward masking. *Journal of Experimental Psychology*, 1966, **71**, 543–549.

KAHNEMAN, D. Method, findings, and theory in studies of visual masking. *Psychological Bulletin*, 1968, **70**, 404–425.

KEIDEL, W. D., KEIDEL, U. O., & WIGAND, M. E. Adaptation: Loss or gain of sensory information? In W. A. Rosenblith (Ed.), *Sensory communication*. New York, Wiley: 1961.

POLLACK, I. Methodological examination of the PEST (Parametric Estimation by Sequential Testing) procedure. *Perception & Psychophysics*, 1968, **3**, 287–289.

POLLACK, I. Methodological examination of the PEST (Parametric Estimation by Sequential Testing) procedure: II. *Perception & Psychophysics*, 1971, **9**, 229–230.

SPERLING, G. Temporal and spatial visual masking: I. Masking by impulse flashes. *Journal of the Optical Society of America*, 1965, **55**, 541–549.

TAYLOR, M. M., & CREELMAN, C. D. PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America*, 1967, **41**, 782–787.

(Received February 25, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 251–257

RECALL AND RESISTANCE TO UNLEARNING OF VERBAL MEDIATING ASSOCIATES AS A FUNCTION OF ANTICIPATION INTERVAL[1]

TERRY H. EBERT[2] AND DANIEL FALLON[3]

State University of New York at Binghamton

J. Richardson's 1968 finding of facilitated learning on a test list by chaining-paradigm over nonchaining-paradigm groups with 4-sec., but not 1-sec., anticipation interval (AI) was replicated. At the end of test-list learning, significantly more mediating associates were recalled by chaining-paradigm than by nonchaining-paradigm groups. Within chaining-paradigm groups, significantly more mediating associates were recalled by the 4-sec. AI group than by the 1-sec. AI group. After an additional 12 trials of test-list learning at a uniform 1-sec. AI, recall by a previously 4-sec. AI chaining-paradigm group was indistinguishable from a previously 1-sec. AI chaining-paradigm group, thereby suggesting unlearning of the mediating associates in the 4-sec. AI group. Resistance to unlearning during these 12 trials was exhibited by a chaining-paradigm group previously exposed to a random mixture of 4-sec. and 1-sec. AIs for the same paired-associate items over successive trials of the test list.

In situations where *S*s are assumed to be engaged in specific covert behavior, such as in mediation or coding paradigms, manipulation of the duration of the recall or anticipation interval (AI) can have markedly greater effects upon performance than in control situations where similar covert behavior is not assumed to occur (e.g., Podd & Spear, 1967; Richardson, 1967). As Richardson (1968) has noted, however, this effect is critically dependent upon the values which are selected for the temporal parameters of study interval and AI.

Richardson (1968) used the B-C-D associative chains of Russell and Storms (1955) in a four-stage transfer experiment with both chaining and nonchaining paradigms. Relative to appropriate nonchaining controls, a chaining-paradigm group with 4-sec. AI performed significantly better on the test list than a chaining-paradigm group with 1-sec. AI. It was reasoned that the 4-sec. group had sufficient AI time in which to use the mediation device covertly, and the 1-sec. group did not, thus accounting for the differential facilitation. The present experiment was designed to confirm and elaborate these findings by measuring directly post-test-list recall of the mediating associates. Although mediators may never appear overtly during learning of the test list, those *S*s who consistently recite the mediators covertly during test-list learning should be able to recall them better than those who do not recite them (cf. Weaver & Schulz, 1968).

It is reasonable to assume for a mediated test list that in both short (e.g., 1-sec.) and long (e.g., 4-sec.) AI conditions mediating associates are elicited covertly, but in the short AI condition lack of time to complete the mediating chain results in either no response or an incorrect response. Thus, the elicitation conditions are favorable for unlearning of the mediating associates to occur (Keppel, 1968) in the short AI condition. Conversely, in the long AI condition elicited covert mediating associates are provided with sufficient time to produce a correct response and should thus be strengthened rather than unlearned. These premises lead to the prediction that recall of the mediating associates will be greater following learning of a mediated test list with long AI than with short AI.

---

[1] This article is based on a thesis submitted to the Graduate School of Arts and Sciences, State University of New York at Binghamton, in partial fulfillment of the requirements for the MA degree by the first author under the supervision of the second author. The experiment grew from ideas originally formulated during a luncheon conversation of the second author with J. Richardson and N. E. Spear, to both of whom this indirect debt is acknowledged, although the authors assume the full burden of responsibility for interpretation, design, and execution.

[2] Now at Department of Educational Administration, Hofstra University, Hempstead, N.Y. 11551.

[3] Requests for reprints should be sent to Daniel Fallon, Department of Psychology, State University of New York, Binghamton, N.Y. 13901.

Further, it may be possible to train *S*s to persist in a strategy of using the mediating associates in the face of a short AI by mixing long and short AI conditions for each item pair over successive trials within a given list. If *S*s are unable to predict whether a given AI will be long or short, but are able to complete a covert response chain only when it is long, the successful recitation of the mediating chain and consequent correct response on trials when the item appears with a long AI may cause persistence of this strategy on trials when the same item appears with a short AI. To the extent that such persistence training leads to high recall of mediating associates, relative to groups without this training, following subsequent performance on a test list under uniformly short AI conditions, the effect may be characterized as resistance to unlearning.

The present experiment sought (*a*) to confirm reports of facilitated learning in a chaining paradigm when sufficient time is provided during AI for mediation to occur, (*b*) to measure recall of the assumed mediators as an indirect test of whether mediation occurred as a function of chaining paradigm and of AI, and (*c*) to impose a condition with a random mixture of items of long AI with items of short AI in a single list for the purpose of producing resistance to unlearning of the mediators.

METHOD

Subjects.—The *S*s were 180 undergraduate male and female Harpur College students, who were required to participate in a certain number of experiments for an introductory course in psychology. The *S*s had not previously participated in an experiment using the Russell and Storms (1955) materials. The *S*s were assigned to conditions using a block randomization procedure which balanced for sex and order of appearance at the laboratory. All *S*s received the same instructions for the first two pretraining lists of the experiment and received specific instructions for their particular groups in later parts of the experiment. All *S*s were tested by the same *E* and were run singly. In addition to the 180 *S*s who completed the experiment, 9 were dropped for various reasons including equipment failure, or failure to learn a list or to follow instructions. Each of the 9 was in a different group.

Materials and apparatus.—Four of five lists consisting of 10 pairs each of verbal stimuli and responses (B-C, C-D, S-B, S-N, S-D) were used in each of two paradigms. The B-C and C-D lists were common word associates taken exactly from the Russell and Storms (1955) materials, e.g., B = stem, C = flower, D = smell, (cf. Richardson, 1968). The S terms were CVCs of low formal similarity (Archer, 1960). The N terms were a set of numerals, 0 through 9. The chaining paradigm consisted of the following lists learned in order: B-C, C-D, S-B, and S-D. The nonchaining paradigm consisted of the following lists learned in order: B-C, C-D, S-N, and S-D.

All PA lists were typed in black ink on vellum tape and were presented to *S* on a Stowe memory drum by the anticipation method. The *E* was hidden from the view of *S* by a wooden shield.

The stimuli for the recall of the mediators were S terms typed in black ink on vellum tape and presented on the Stowe memory drum.

Procedure and design.—All groups first learned List B-C at a 3:1 rate to a criterion of one perfect trial, and then learned List C-D, also at a 3:1 rate, to the same criterion. The *S*s in the chaining paradigm (groups labeled C) then learned List S-B at a 2:2 rate to a criterion of two successive perfect trials, while *S*s in the nonchaining paradigm (groups labeled NC) learned List S-N, also at 2:2 rate, to this criterion.

Ten S-D trials were given to all groups. In all groups, the first 3 S-D trials were always at a 4:1 rate. Over Trials 4–10, one-third of the chaining and one-third of the nonchaining groups learned S-D at a 4:1 rate (groups labeled C-4, NC-4), one-third learned S-D at a 1:1 rate (groups labeled C-1, NC-1), and the final third learned S-D at a composite rate containing a mixture of items presented at 4:1 and 1:1 rates (groups labeled C-M, NC-M). For each *S* in Groups C-M and NC-M, all items appeared at both 4:1 and 1:1 rates. For each *S*, the number of 4:1 pairs on any given trial in Groups C-M and NC-M was varied from 0 to 10. The order in which trials containing varying numbers of 4:1 pairs were presented was arranged to form counterbalanced sequences between *S*s so that averaged across all *S*s on a given trial, 50% of the pairs were 4:1 pairs. Therefore, for each of Groups C-M and NC-M taken across all *S*s there was an equal, 50%, proportion of 4:1 and 1.1 pairs on each of Trials 4–10.

In both the chaining and the nonchaining conditions, groups with AI schedules of either 1 sec. or mixed seconds were asked to solve a pyramid puzzle after S-D learning. This was done so as to equalize the amount of time from the beginning of S-D learning to the recall of the B-C mediators in all groups. Groups C-1 and NC-1 were given the puzzle for 3.5 min., and Groups C-M and NC-M were given the same puzzle for 1.75 min.

When all groups had finished 10 trials of S-D learning, half of each of the groups was tested for verbal recall of B-C by a nonpaced, cued procedure; i.e., *S* was presented the S terms one at a time, and they remained in view of *S* until either a response was given or *S* reported that he did not know the response. The *S* was permitted to take as much time to respond as desired. After a response was made the drum was advanced so that the next *S* term was presented. The recall had to be given in the correct, i.e., B-C, order, and both members had to be given for the recall response to be scored as a correct response. The *S* was not shown the response terms and was not told if his responses were correct.

All groups were then presented with an additional 12 S-D learning trials at a 1:1 rate. For the half of the groups which did not receive the first recall procedure, this "fast" S-D learning occurred 1 min. following regular S-D learning. After the "fast" S-D learning, all 12 groups were tested for recall of B-C using the cued, nonpaced procedure described above.

The intertrial interval was 4 sec. for the B-C, C-D, S-B, and S-N lists. It was 5 sec. for the S-D lists. There was a rest period of 1 min. between each of the learning tasks of the experiment. A blank yellow tape was presented to *S* during each rest period. The items within each of the PA lists during each of the learning tasks as well as the items in the fast learning task were arranged in five random orders. The order of the S terms in the recall task was also randomized.

In summary, three experimental variables were manipulated: (*a*) Paradigm had 2 levels; chaining and nonchaining; (*b*) AI had 3 levels; 4 sec., 1 sec., and a specially constructed composite mixed rate of 4 sec. and 1 sec; (*c*) mediator-recall test had 2 levels; prior to, and following, a fast-rate learning task. Thus, the experimental design was a 2 × 3 × 2 factorial. There were 15 *S*s in each of the 12 groups.

RESULTS

The 12 groups of *S*s did not differ on the number of trials to learn the first, B-C, list ($\bar{X} = 2.84$; range of $\bar{X}$s $= 2.67$–3.20), or on the number of trials to learn the second, C-D, list ($\bar{X} = 3.42$; range of $\bar{X}$s $= 3.13$–3.80). On the third list, the 6 nonchaining paradigm groups learning List S-N took significantly fewer trials to criterion ($\bar{X} = 11.60$) than the 6 chaining-paradigm groups learning List S-B ($\bar{X} = 14.28$), F (1,168) $= 9.27$, $p < .005$. It was assumed that List S-N was easier to learn than List S-B because the numerals reduced response-learning requirements (Battig, 1962), facilitated associative learning, and increased list differentiation.

Fourth-List, S-D, Learning

The mean number of correct responses for each of the first 10 trials of learning with List S-D is presented in Fig. 1. For the first 3 trials of S-D learning, all *S*s learned the list at a uniform AI of 4 sec.; i.e., the different AI schedules were introduced first on Trial 4. Therefore, the data were analyzed separately for the first 3 S-D learning trials and the last 7 S-D learning trials.

Trials 1–3.—As Fig. 1 clearly shows, paradigm had a significant effect upon performance during the early trials, with *S*s in the chaining paradigm performing at a higher level than *S*s in the nonchaining paradigm, F (1, 168) $= 378.10$, $p < .001$. Furthermore, the differential in performance due to paradigm increased over the three successive trials, F (2, 336) $= 10.74$, $p < .001$.

Transition of Trial 3 to Trial 4.—An analysis of variance on the data of Trials 3 and 4 indicated a significant AI × Paradigm × Trial interaction, F (2, 174) $= 7.36$, $p < .01$. The immediate effect of variations in AI on Trial 4 was indicated by the AI × Trial interaction, F (2, 174) $= 160.90$, $p < .001$.

Data describing the change in performance from Trial 3 to 4 are presented in Table 1. Because all *S*s were exposed to an AI of 4 sec. through Trial 3, the only effect evident upon performance on Trial 3 was the main effect of paradigm. Thus, for example, on Trial 3 there was a significant difference in the performance of Groups C-1 and NC-1, which could be attributed to paradigm. However, when AI was dropped to 1 sec. for these two groups on Trial 4, the significant effect of paradigm disappeared. However, for Groups C-4 and NC-4, which were maintained at an AI of 4 sec., significant differences in performance evident on Trial 3 remained on Trial 4.

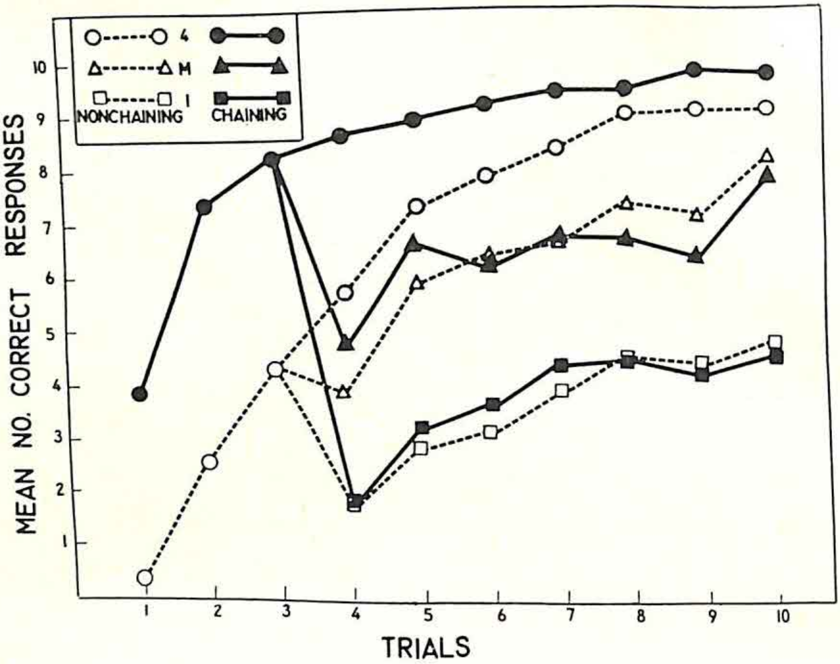


FIG. 1. Mean number of correct responses as a function of trials, with paradigm and AI as parameters. During Trials 1–3 AI was 4 sec. for all *S*s.

Trials 4–10.—During the next seven S-D learning trials, AI continued to affect performance, F (2, 168) = 171.92, $p <$ .001. Because of the continuing strong effect of paradigm on the 4-sec. AI groups, the overall main effect of paradigm was also significant during these trials, F (1, 168) = 4.08, $p < .05$. A significant AI × Paradigm interaction, F (2, 168) = 3.24, $p < .05$, confirmed that AI produced more pronounced effects upon performance within the context of the chaining paradigm. There was greater improvement over trials in the 4-sec. AI groups than in the mixed-second AI groups than in the 1-sec. AI groups, F (12, 1008) = 2.58, $p < .01$. There was also a greater improvement in performance over trials among the nonchaining-paradigm groups than among the chaining-paradigm groups, F (6, 1008) = 6.68, $p < .001$.

TABLE 1

CONTRAST OF MEAN NUMBER OF CORRECT RESPONSES ON TRIALS 3 AND 4 OF THE S-D LIST AS A FUNCTION OF PARADIGM AND ANTICIPATION INTERVAL

Group	Trial 3			Trial 4		
	$\bar{X}$	t (58)	p	$\bar{X}$	t (58)	p
C-4 NC-4	8.50 3.83	11.12	.001	8.70 5.77	6.21	.001
C-M NC-M	8.33 4.93	6.73	.001	4.70 3.80	2.28	.05
C-1 NC-1	8.10 4.30	8.17	.001	1.73 1.63	.27	*ns*

Trials 11–22.—After half of the *S*s were given a test for recall of the mediating associates, all *S*s were presented with 12

TABLE 2

MEAN NUMBER OF CORRECT RESPONSES ON THE S-D LIST DURING TRIALS 11–22

AI during Trials 4–10	Chaining paradigm		$\bar{X}$	Nonchaining paradigm		$\bar{X}$
	Recall 1	Recall 2		Recall 1	Recall 2	
4 sec.	6.32	6.79	6.55	6.76	6.76	6.76
Mixed sec.	5.40	7.10	6.25	7.54	7.92	7.73
1 sec.	4.99	6.28	5.64	6.10	6.08	6.09
$\bar{X}$	5.57	6.72	6.18	6.80	6.92	6.86

Note.—Anticipation interval (AI) = 1 sec. on all trials.

additional learning trials on List S-D at a uniform 1-sec. AI. A summary of these data is presented in Table 2. Performance of *S*s in the nonchaining paradigm was superior to that of *S*s in the chaining paradigm during Trials 11–22, F (1, 168) = 5.22, $p < .05$. Those *S*s who had been exposed to the recall test between Trials 10 and 11 (Recall 1) exhibited poorer performance than those who did not receive the recall test until after Trial 22 (Recall 2), F (1, 168) = 4.19, $p < .05$. There were residual differences in performance due to differences in response strength produced by AI during Trials 4–10, F (2, 168) = 4.59, $p < .05$. However, the comparison of the 1-sec. AI groups with the other groups accounted for 91% of the sum of squares for the main effect of AI, F (1, 168) = 8.39, $p < .005$, indicating that the low initial response strength of the 1-sec. AI groups was principally responsible for the effect.

Recall of the Verbal Mediating Associates

The mean number of correct responses recalled at the conclusion of two stages of learning with List S-D is presented in Table 3. Recall 1 indicates recall of the mediators for the six groups which were first tested after 10 trials of S-D learning, and Recall 2 indicates recall of the mediators for the six groups which were first tested after the complete 22 trials of S-D learning. Because the within-*S* data of those *S*s tested for recall of the mediators both at Recall 1 and Recall 2 showed a pattern similar to that of the independent groups and provided no additionally useful information, only the data for the independent groups are presented.

TABLE 3

MEAN NUMBER OF CORRECT B-C MEDIATORS RECALLED PRIOR TO TRIALS 11–22 OF S-D LEARNING (RECALL 1) AND FOLLOWING TRIALS 11–22 (RECALL 2)

AI during Trials 4–10	Chaining paradigm		Nonchaining paradigm	
	Recall 1	Recall 2	Recall 1	Recall 2
4 sec.	6.53	4.53	1.73	2.53
Mixed sec.	5.20	6.27	1.80	3.60
1 sec.	4.07	5.07	1.33	1.73

Note.—Anticipation interval during Trials 11–22 = 1 sec.

The *S*s in the chaining-paradigm groups recalled significantly more mediators than *S*s in the nonchaining-paradigm groups, F (1, 168) = 55.34, $p < .001$. A separate analysis on the data for the nonchaining-paradigm groups alone revealed no effects which approached significance. For the chaining-paradigm groups taken alone, the predicted AI × Recall Test interaction reached a marginal level of significance, F (2, 168) = 3.04, $p < .06$. Further, 99.96% of the interaction sum of squares was accounted for by the comparison over the two recall tests of Group C-4 with the other two groups, F (1, 168) = 6.13, $p < .025$. At Recall 1, Group C-4 differed significantly from Group C-1, t (28) = 2.22, $p < .05$, but the two groups did not differ significantly at Recall 2, $t(28) = .81$. At Recall 2, the prior prediction that the "persistence training" for Group C-M during Trials 4-10 would result in higher recall than for Groups C-4 and C-1 was confirmed, t (43) = 1.87, $p < .05$, one-tailed.

DISCUSSION

The differential facilitation of test-list learning in the chaining paradigm under a 4-sec. AI replicates Richardson (1968) and is consonant with the notion that because mediation takes time, facilitation can only occur when AI is sufficiently long to permit it. Of interest in the present experiment was the demonstration of the immediate direct effect of AI: the performance advantage of a chaining-paradigm group over a nonchaining-paradigm group was lost completely when the AI was dropped from 4-sec. to 1-sec. on Trial 4. This finding emphasizes the dependence of chaining-paradigm superiority upon AI and, by implication, upon sufficient time for the recitation of a mediation device.

The use of a recall test as a measure of degree of mediation is a simple extension of the procedures developed by Barnes and Underwood (1959) and successfully extended to different situations by others (e.g., Weaver & Schulz, 1968). Among the distinguishing features of the present procedure are that recall of two mediating associates in sequence was required and that recall was measured as a function of AI in the mediated-learning task. The results of the present experiment demonstrate that more mediators are recalled following learning with a 4-sec. AI than with a 1-sec. AI. Further, there is a suggestion in the data that available mediators are unlearned when AI in the mediated-learning task is uniformly 1-sec. and that *S*s subjected to a mixture of 1-sec. and 4-sec. AIs during mediated learning are resistant to this unlearning.

The nature of the learning task was such that *S*s in the nonchaining paradigm could recall the correct mediating associates only by first recalling a D term and then reconstructing the B and C terms from it, i.e., in the sequence S-D-C-B. For these *S*s, recall of the third list, S-N, was of no benefit in recalling the B-C mediators. However, *S*s in the chaining paradigm could recall the mediators by recalling the third list, S-B, and then retrieving the mediating associates, i.e., in the sequence S-B-C-D. Thus, complete unlearning of the S-B connection for chaining-paradigm *S*s would be indicated by recall performance similar to nonchaining-paradigm *S*s. The finding that recall was substantially better for the 1-sec. chaining-paradigm groups than for the 1-sec. nonchaining-paradigm groups argues, therefore, that the mediating associates were not completely unlearned in the chaining paradigm as a result of the 1-sec. AI, even though performance was not facilitated during test-list learning.

There are several shreds of evidence that *S*s in the chaining paradigm were persisting in a strategy of mediation, even when the strategy did not benefit performance. For example, the poorer performance of the chaining-paradigm groups during "fast" test-list learning, Trials 11–22, suggests that these *S*s, taken together, were persisting in a mediation strategy which was actually detrimental under the new circumstances. Similarly, the recall test administered between Trials 10 and 11 could be expected, by providing an occasion for overt recitation and rehearsal, to strengthen the mediating associates. Thus, the poorer performance of *S*s during "fast" test-list learning, Trials 11–22, who were given the recall test prior to Trial 11 is suggestive of persistence in an inappropriate mediation strategy.

The outcome of Recall Test 2, in which the superior recall of Group C-4 over Group C-1 at Recall Test 1 was eliminated, implies that persistence in a mediation strategy led to the unlearning of mediators for Group C-4. Because there was no control group which continued to learn List S-D with a 4-sec. AI during Trials 11–22, it is not possible to rule out definitively the possibility that the loss over the two tests for Group C-4 was due to mediator "drop-out" (cf. Weaver & Shulz, 1968). However, the complete absence of any loss for Group C-M argues against this possibility. In fact, the reversal in relative superiority of recall of Groups C-M and C-4 over the two recall tests is supportive of the proposition that *S*s in Group C-M exhibited resistance to unlearning of the mediators.

It is possible to conceptualize the conditions for Group C-M during Trials 4–10 as a partial reinforcement situation if one assumes, unlike the usual paired-associate reinforcement model (e.g., Schulz & Runquist, 1960; Silverstein, 1967), that a correct response must be emitted overtly, or at least covertly elicited, in order to be reinforced. Thus, a 4-sec. AI provided sufficient time for elicitation of the mediating chain and of the correct response and resulted in reinforcement. However, a 1-sec. AI permitted elicitation of the initial portions of the mediating chain, but cut off the chain after 1 sec., thereby blocking elicitation of the correct response and resulting in nonreinforcement. The mixture of these two events in unpredictable fashion trained *S*s in Group C-M to persist in a mediation strategy under unfavorable conditions, and thus led to resistance to unlearning of the mediators. This argument may be viewed as a direct analogy to persistence theory developed from simple instrumental learning situations (e.g., Amsel, 1958–1972).

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin*, 1958, **55**, 102–119.

AMSEL, A. Behavioral habituation, counterconditioning, and a general theory of persistence. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II*. New York: Appleton-Century-Crofts, 1972.

ARCHER, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, **74** (10, Whole No. 497).

BARNES, J. M. & UNDERWOOD, B. J. "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology*, 1959, **58**, 97–105.

BATTIG, W. F. Paired-associate learning under simultaneous repetition and nonrepetition conditions. *Journal of Experimental Psychology*, 1962, **64**, 87–93.

KEPPEL, G. Retroactive and proactive inhibition. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

PODD, M. H., & SPEAR, N. E. Stimulus relatedness and response coding. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 55–60.

RICHARDSON, J. The locus of mediation and the duration of the anticipation interval in a transfer paradigm. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 247–256.

RICHARDSON, J. Latencies of implicit associative responses and positive transfer in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 638–646.

RUSSELL, W. A., & STORMS, L. H. Implicit verbal chaining in paired-associate learning. *Journal of Experimental Psychology*, 1955, **49**, 287–293.

SCHULZ, R. W., & RUNQUIST, W. N. Learning and retention of paired adjectives as a function of % ORM. *Joruna1 of Experimental Psychology* 1960, **59**, 409–413.

SILVERSTEIN, A. Unlearning, spontaneous recovery, and the partial reinforcement effect in paired-associate learning. *Journal of Experimental Psychology*, 1967, **73**, 15–21.

WEAVER, G. A., & SCHULZ, R. W. A-B, B-C, A-C mediation paradigm: Recall of A-B following varying numbers of trials of A-C learning. *Journal of Experimental Psychology*, 1968, **78**, 113–119.

(Received November 3, 1971)

urnal of Experimental Psychology
72, Vol. 95, No. 2, 258–263

MOTION PARALLAX AND ABSOLUTE DISTANCE[1]

STEVEN H. FERRIS[2]

Naval Submarine Medical Research Laboratory, Groton, Connecticut

The accuracy of absolute distance estimation based on monocular motion parallax was determined both before and after specific training. With the usual distance information eliminated, *S*s either held their heads stationary or rhythmically rotated their heads about a vertical axis while judging the distance of stimuli placed 1.22–4.57 m. away. Although distance perception was poor before training, head movement produced more accurate judgments than head fixed. After only 10 training trials, accurate judgments based on motion parallax were obtained. Results with a white background were as good as with a textured background when *S*s were given direct information about motion parallax. Good results were also obtained for motion parallax relative to a near reference object, 76.2 cm. away. The results indicate that motion parallax can be useful for absolute distance estimation.

Motion parallax has been defined by Graham (1965, p. 504) as follows: "When a subject's eyes move with respect to the environment, or when the environment moves with respect to the subject's eyes, a differential angular velocity exists between the line of sight to a fixated object and the line of sight to any other object in the visual field." If *O* rotates his head about a vertical axis, objects more distant than the fixated object appear to move in the same direction as the head movement, at velocities dependent on their distances from *O*; similarly, objects closer than the fixated object move at velocities dependent on their distances, but they move in the opposite direction. Although Helmholtz (1962, pp. 295–297) believed that motion parallax information is useful for perceiving both relative depth (distance between objects) and absolute depth (distance between object and *O*), most of the relevant experiments have been concerned only with relative depth (Gibson, Gibson, Smith, & Flock, 1959; Graham, Baker, Hecht, & Lloyd, 1948; Smith & Smith, 1963; Tschermak-Seysenegg, 1939, cited in Gibson et al., 1959, p. 41). There seems to be no doubt that motion parallax is a useful (if not precise) source of information for perceiving relative depth. Even infants and newborn animals apparently use motion parallax (cf. Bower, 1965; Walk & Gibson, 1961).

A limitation of many of the motion parallax studies is that they were concerned with visual field movement in the absence of *O* movement. Park (1964) has shown that visual field movement alone, passive *O* movement, and active *O* movement, in that order, are progressively more effective for relative depth perception. Park also presented an analysis indicating that motion parallax cannot provide a basis for perceiving absolute depth. His negative experimental results agree with those of Bourdon (1902, cited in Gibson et al., 1959, p. 41). But Dees (1966), in the only extensive attempt to use motion parallax for absolute distance, was able to train *S*s to make accurate absolute distance estimates based solely on motion parallax. Unfortunately, Dees did not use actual *O* movement, and the performance of his *S*s before training was not reported. The present experiments were designed to reexamine the use of motion parallax for absolute distance. The *S*s actively moved their heads from side to side, and results before and after training are reported.

---

[1] This research was supported by the Bureau of Medicine and Surgery, United States Navy Department, Research Work Unit M4306.03-2050DXC5. The opinions or assertions contained herein are the private ones of the writer and are not to be construed as official or reflecting the views of the Navy Department, the Naval Submarine Medical Research Laboratory, or the Naval Service at large.

[2] Requests for reprints should be sent to Steven H. Ferris, Vision Branch, Naval Submarine Medical Research Laboratory, Naval Submarine Medical Center, Naval Submarine Base New London, Groton, Connecticut 06340. This research was conducted while the author held a National Research Council-Bureau of Medicine and Surgery Postdoctoral Research Associateship. I thank Jo Ann Kinney for her valuable advice.

General Method

Subjects.—A total of 70 Navy enlisted men served as *S*s. Each had normal vision and took part in only one experiment. None had any prior knowledge of the experimental situation.

Apparatus.—The experimental arrangement permitted presentation of targets at various distances, with minimal distance information available. The *S* sat in a large room, facing a wall which was 5.92 m. away. The field of view was restricted by a black screen placed 41.91 cm. from *S*'s eye, and having a rectangular opening 73.66 cm. wide × 12.70 cm. high (82°36′ × 17°14′). Neither the floor nor the ceiling was visible. Two curved partitions, draped with white sheets, were placed between the wall and each side of the screen, thus enclosing the area between the screen and the wall. Depending on the experimental condition, the wall was covered with either solid white paper or white paper having a texture of 1.27-cm. red squares (290 squares/sq. m.). These backgrounds were 3.35 m. wide (67°54′), had an effective height of 17°14′, and were illuminated to a brightness of 25 ftl. Since this was the only illumination in the room, the targets, which were black, appeared as silhouettes. The targets varied in width from 4.44–19.05 cm., and when individually presented at the appropriate distances, a constant retinal width of 2°22′ was maintained. The effective height was always 17°14′. The basic test distances were 4, 5, 7, 9, 12, and 15 ft. (1.22, 1.52, 2.13, 2.74, 3.66, and 4.57 m.). Other distances were included during training sessions. All viewing was monocular.

Procedure.—There were three phases to each experiment: The initial test (T_1), 10 training trials, and the final test (T_2). Each test distance was presented twice, in random order. The *S*'s task was to judge the absolute distance of the target, to the nearest foot or half-foot. Ten different distances were presented during training. Training consisted of *S* being verbally informed of the correct distance after each judgment. Between each judgment, the viewing opening was covered while a new target was positioned. The major experimental variable was the presence or absence of motion parallax information. During both test and training judgments, *S* either kept his head fixed on a chin rest, or rotated his head rhythmically from side to side about a vertical axis as he viewed the target.

In the instructions, *S*s were asked to look directly at the target and report how far from them it appeared to be. It was stressed that the task was purposely made difficult and that they should not be concerned if they had some difficulty.

The two judgments of each *S* at each distance were averaged, and the median judgments for each group were plotted as a function of physical distance. In addition, power-function exponents (Stevens, 1957) were computed.

Experiment I

Method.—The variables examined were textured versus white background, and head fixed versus motion parallax. Ten *S*s were assigned to each of the four treatment combinations. After T_1, *S*s in the motion parallax groups were informed of the nature of motion parallax and were encouraged to use this information in determining their judgments.

Results.—The median distance estimates are plotted in Fig. 1. The semi-interquartile ranges are listed in Table 1 to indicate the variability of the estimates. An analysis of variance for a four-factor design with repeated measures on two factors (Winer, 1962) indicated a statistically significant effect on the distance estimates for physical distance, $F\ (5,\ 180) = 164.7$, $p < .001$, for motion parallax (parallax versus head fixed), $F\ (1,\ 36) = 5.83$, $p < .05$, and for training (T_1 vs. T_2), $F\ (1,\ 36) = 78.79$, $p < .001$, but no significant effect for background (texture

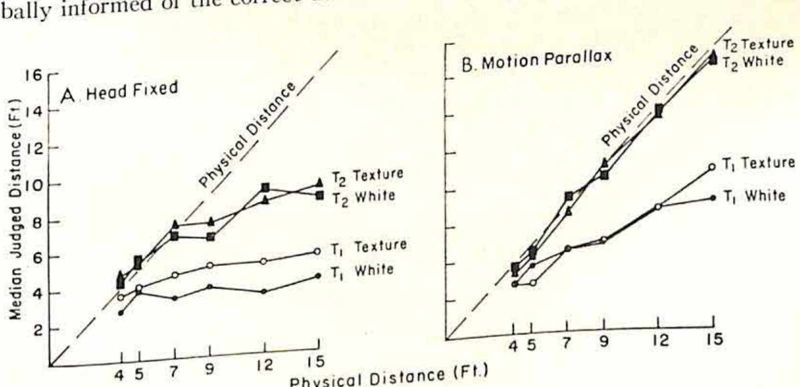


FIG. 1. Median distance estimates for T_1 and T_2 in Exp. I: Textured versus white background, head fixed (A) and motion parallax (B).

TABLE 1

Variability of Distance Judgments (Semi-interquartile Ranges)

Group	Measure (in ft.)											
	4		5		7		9		12		15	
	T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2
Exp. I												
White-fixed	1.50	.75	2.75	1.50	2.25	1.00	2.00	1.00	1.25	2.25	1.25	3.25
White-parallax	.75	.67	1.25	.67	.75	1.50	.75	2.62	1.37	1.87	1.62	1.75
Texture-fixed	.62	1.00	.50	1.25	.87	1.25	.87	.62	.50	1.12	1.00	1.75
Texture-parallax	1.00	.75	1.00	.25	1.00	.50	.50	.37	1.25	1.50	2.25	.75
Exp. II												
White-uninformed	1.00	.50	.75	.37	1.50	1.12	1.62	.62	2.00	1.50	3.00	.50
Texture-uninformed	.50	.37	.62	.87	.62	.62	.62	1.25	1.75	.50	1.00	1.75
Exp. III												
Near-fixed	.75	—	1.37	—	1.00	—	1.37	—	1.37	—	2.25	—
Near-parallax	—	.37	—	.75	—	.62	—	.87	—	1.75	—	.50

versus white), $F\ (1,\ 36) < 1.0$. Defining judgment accuracy as closeness to physical distance, comparison of Fig. 1A and 1B indicates that both before and after training, judgments were more accurate for parallax than for head fixed. Furthermore, the advantage of head movement increased with distance, as indicated by a significant Parallax × Distance interaction, $F\ (5,\ 180) = 36.12$, $p < .001$. Another way of interpreting this interaction is that the degree of increase in judgment magnitude with increased distance was greater for parallax than for head fixed. There was also a significant Training × Distance interaction, $F\ (5,\ 180) = 34.49$, $p < .001$, reflecting a greater improvement with increased distance. Finally, there was a significant Parallax × Training × Distance interaction, $F\ (5,\ 180) = 3.11$, $p < .01$, indicating that the effect of training increased with distance to a greater extent for parallax than for head fixed. With the exception of a small Background × Distance interaction, $F\ (5,\ 180) = 2.48$, $p < .05$, which had no meaningful interpretation, all other interaction effects were nonsignificant ($p\text{s} > .10$).

Since distance estimation was more accurate for parallax than for head fixed even before *S*s were informed of the nature of motion parallax, it is of interest to know whether *S*s were actually aware of the parallax information. When the parallax *S*s were questioned after T_1, only one of the 10 *S*s in each background group reported that he noticed relative movement between target and background.

Power-function exponents for the T_1 and T_2 distance functions of the four groups are listed in Table 2. The coefficients of determination (r^2) are also listed, to indicate how well the data fit power functions. An exponent of 1.0 would indicate a linear relationship between judged and physical distance, equivalent to perfect distance discrimination. An exponent of 1.0 does not necessarily mean that the judgments were accurate, that is, equivalent to physical distance, since there can still be a large constant error. Taking closeness of the exponent to 1.0 as the measure of discrimi-

TABLE 2

Power-Function Exponents

Group	T_1		T_2	
	Exponent	r^2	Exponent	r^2
Exp. I				
White-fixed	.19	.43	.52	.90
White-parallax	.65	.95	1.04	.99
Texture-fixed	.32	.96	.51	.94
Texture-parallax	.88	.97	1.11	.99
Exp. II				
White-uninformed	.85	.97	.93	.99
Texture-uninformed	.80	.99	.99	.99
Exp. III				
Near-fixed	.45	.94	—	—
Near-parallax	—	—	1.03	.99

nation, it is apparent from Table 2 that discrimination was quite poor before training for head fixed, but much better for parallax. Discrimination improved as a result of training.

Discussion.—The very poor distance discrimination with head fixed before training indicates that the experimental arrangement successfully eliminated virtually all distance information. The slight increase in the estimates at the nearer distances was probably due to accommodation and accommodative convergence. A particular *S* tended to perceive the target at a particular distance, usually only a few feet away, regardless of the physical distance. As Gogel (1969) has pointed out, there is a "specific distance tendency" which, in the absence of distance information, results in objects being perceived at a definite near distance. Since the nearest physical distance of the target was 4 ft. (1.22 m.), this tendency accounts for the fact that the judgments were almost always less than the physical distance.

The very low power-function exponents before training are consistent with the results for "reduced cues" obtained by Künnapas (1968). Training with head fixed led to some improvement, indicating that some residual distance information remained. Besides comments such as "how it looked," some *S*s reported they based their judgments on size differences. Even though retinal size was constant, the more distant, physically wider targets appeared wider. This change in apparent size for targets of equal retinal angle is quite surprising. There was virtually no change in perceived distance even though the changes in perceived size imply that some distance information was available. A speculative explanation for this outcome is that distance and size perception involve parallel processes. If this were true, it might be possible for distance information sufficient for producing changes in apparent size to be insufficient for producing changes in perceived distance.

An unexpected result of the experiment was that the presence or absence of a textured background had no important effect on the usefulness of motion parallax information. It is especially surprising that background had no effect after training. A possible explanation for this result is that once *S* was looking for relative movement, there was enough noticeable texture on the white background to permit maximal performance. Hence adding additional texture would have no important effect. This possibility was tested in the next experiment.

EXPERIMENT II

Method.—Two additional motion parallax groups were tested, one for each background. There were 10 *S*s in each group. In contrast to the previous experiment, these *S*s were not informed about the nature of motion parallax prior to the training trials. The training procedure was otherwise the same as in the previous experiment.

Results.—The median distance estimates are plotted in Fig. 2. The results show that the presence of a textured background did make a difference after training when *S*s were not told about motion parallax. An analysis of variance for a three-factor design with repeated measures on two factors (Winer, 1962) indicated statistically significant effects for training, F (1, 18) $= 41.5$, $p < .001$, for physical distance, F (5. 90) $= 196.4$, $p < .001$, and for the Training × Distance interaction, F (5, 90) $= 16.2$, $p < .001$. As in Exp. I, training improved performance, judgment magnitudes increased with distance, and the effect of training increased with distance. Overall, background (texture vs. white) did not have a significant effect, F (1, 18) $= 2.0$, $p > .10$, but there was a significant Background × Training interaction, F (1, 18) $= 7.6$, $p < .05$. It is apparent that the distance judgments were more accurate for texture than for white after training, but not before training. There was also a

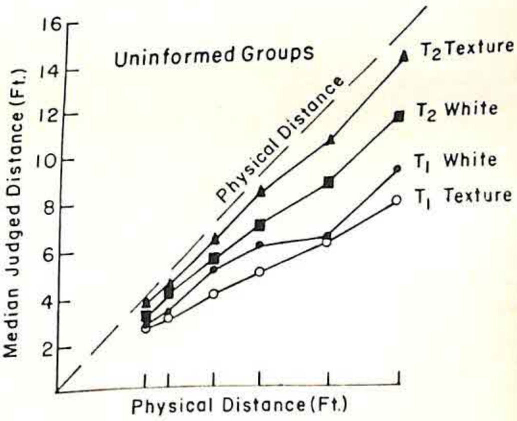


FIG. 2. Median distance estimates, based on motion parallax, for T_1 and T_2 in Exp. II: uninformed groups, textured versus white background.

significant Background × Training × Distance interaction, $F\ (5, 90) = 3.0,\ p < .05$, reflecting the fact that the increased effect of training with increased distance was greater for texture than for white. The power-function exponents for Exp. II (see Table 2) indicate a greater improvement in discrimination for texture than for white. When questioned after the experiment, three *S*s in the texture group and no *S*s in the white group reported that they used motion parallax information.

A comparison of Fig. 2 and 1B indicates that except for T_2-white, the results for this experiment and for the Parallax groups of Exp. I appear to be quite similar. An analysis of variance was performed using the T_1 data from Exp. II and the T_1 data from the parallax groups of Exp. I. Only physical distance was a significant factor, $F\ (5, 180) = 120.9,\ p < .001$, and there were no significant interaction effects (ps $> .10$). Since *S*s of the two experiments were treated exactly the same prior to training, Exp. II therefore replicates the T_1-parallax results of Exp. I. For T_2, however, a separate analysis of variance indicated that the effect of informing *S*s about motion parallax was statistically significant, $F\ (1, 36) = 6.1,\ p < .05$. There was also a significant Informed × Background interaction, $F\ (1, 36) = 5.3,\ p < .05$, indicating that informing *S*s made a difference only for the white background.

EXPERIMENT III

Method.—In the first two experiments, the parallax movement was relative to a background that was more distant than the target. This experiment examined motion parallax relative to an object nearer than the target. A 1.90-cm. wide, vertical black strip (1°26′) was mounted behind the viewing opening, 76.20 cm. from *S* and 3°56′ to the left of the target. The white background was used. A group of 10 *S*s was tested with head fixed (T_1), informed about motion parallax (relative to the near strip) and trained with head movement, and then retested with head movement (T_2).

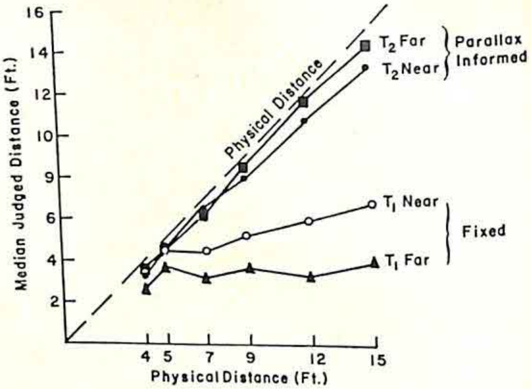


FIG. 3. Median distance estimates for T_1 (head fixed) and T_2 (motion parallax, informed) based on a near reference object (Exp. III) and a far reference background (Exp. I, white background).

Results.—The data demonstrate the effectiveness of motion parallax with a near reference. In Fig. 3, the results are compared with the corresponding results from Exp. I. With head fixed, distance perception with the near reference strip was initially poor, but somewhat better than when it was not present. For T_1, a statistical comparison of the Near and Far conditions revealed a significant Near–Far × Distance interaction, $F\ (5, 90) = 4.4$, $p < .01$, reflecting a greater increase in judgment magnitude with increased distance for near than for far. Also, there was a larger T_1 exponent for near (see Table 2), indicating better initial discrimination when the strip was present. For parallax after training, however, a separate analysis indicated no significant differences between Near and Far (Fs < 1.0). Thus, training with head movement and the near object led to the same improvement as when the far background was used.

GENERAL DISCUSSION

These experiments have demonstrated that prior to any training, absolute distance perception is better with head movement than with head fixed. This result is inconsistent with the view of Park (1964) that motion parallax is not useful for perceiving absolute distance. The probable reason for the negative results of previous investigators is that judgments for a wide range of distances were not obtained. It is significant that the Head Movement condition was superior even though almost all *S*s reported that they were not aware of the motion parallax information. The visual system is apparently capable of using, to some extent, information added when the head is rotated. Since the results before training were similar for the two background conditions, a strongly textured background is not essential for providing the relevant information.

The experiments also demonstrate that training *S*s to use motion parallax information can lead to accurate distance estimation. This result extends the work of Dees (1966) to include the situation in which the visual field movement results from active head movement. It is of interest that whereas Dees' *S*s achieved maximal performance after four presentations of each test distance, *S*s in the present experiments performed well after only 10 trials in which no distance was presented more than once. The rapid learning in these experiments possibly reflects an advantage of active head movement over visual field movement with head fixed. It should be emphasized that the improvement in judgment accuracy after training does not necessarily imply that the *perception* of distance actually improved. It is likely, as confirmed by the questioning of several *S*s, that *S*s merely learned to relate particular amounts of motion parallax to particular physical distances.

Another result of interest concerns the importance of a textured background and of specifically informing *S*s about motion parallax. When directly informed prior to training, *S*s performed similarly regardless of the background. Apparently, only minimal differentiation within the background is necessary when *S*s are consciously responding to motion parallax information. On the other hand, a textured background led to more accurate performance when *S*s were not informed of the nature of motion parallax.

There is an important practical limitation to training *S*s to use motion parallax based on motion relative to a background. Unless the background is always located at infinity, as was the case in Dees' (1966) experiment, or unless the finite distance of the background is permanently fixed, as in Exp. I and II, the proper scaling of motion parallax as a function of stimulus distance will vary with differences in the distance of the background. Consequently, training in the use of motion parallax with a particular background distance is probably of limited practical value. This limitation was the reason for examining the use of a near reference object in Exp. III. An *S* can always place a near object, such as an outstretched finger, in front of him. Since *S*s successfully used motion parallax with a near reference, the practical limitation of depending on a background can be eliminated.

REFERENCES

Bower, T. G. R. Stimulus variables determining space perception in infants. *Science*, 1965, **149**, 88–89.

Dees, J. W. Accuracy of absolute visual distance and size estimation in space as a function of stereopsis and motion parallax. *Journal of Experimental Psychology*, 1966, **72**, 466–476.

Gibson, E. J., Gibson, J. J., Smith, O. W., & Flock, H. Motion parallax as a determinant of perceived depth. *Journal of Experimental Psychology*, 1959, **58**, 40–51.

Gogel, W. C. The sensing of retinal size. *Vision Research*, 1969, **9**, 1079–1094.

Graham, C. H. Visual space perception. In C. H. Graham (Ed.), *Vision and visual perception.* New York: Wiley, 1965.

Graham, C. H., Baker, K. E., Hecht, M., & Lloyd, V. V. Factors influencing thresholds for monocular movement parallax. *Journal of Experimental Psychology*, 1948, **38**, 205–223.

Helmholtz, H. von. *Physiological optics.* Vol. III. New York: Dover, 1962.

Künnapas, T. Distance perception as a function of available visual cues. *Journal of Experimental Psychology*, 1968, **77**, 523–529.

Park, J. N. *The role of motion parallax in depth perception.* (Doctoral dissertation, University of Kansas) Ann Arbor, Mich.: University Microfilms, 1964, No. 65-7668.

Smith, O. W., & Smith, P. C. On motion parallax and perceived depth. *Journal of Experimental Psychology*, 1963, **65**, 107–108.

Stevens, S. S. On the psychophysical law. *Psychological Review*, 1957, **64**, 153–181.

Walk, R. D., & Gibson, E. J. A comparative and analytical study of visual depth perception. *Psychological Monographs*, 1961, 75 (15, Whole No. 519).

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received November 12, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 264–274

INTERFERENCE IN SHORT-TERM MOTOR MEMORY: INTERPOLATED TASK DIFFICULTY, SIMILARITY, OR ACTIVITY?[1]

BARRY H. KANTOWITZ[2]

Purdue University

Four experiments were conducted to evaluate effects of motor interpolated tasks upon retention on a discrete motor movement guided by both kinesthetic and visual-plus-kinesthetic cues. When a paced information conservation interpolated task with two levels of information transmitted per unit time (2.3 and 10 bits/sec) was used, no effect of interpolated task difficulty was found for either kinesthetic or visual modalities with either absolute or algebraic error as the dependent variable (Exp. I and II). When the relative similarity of interpolated and retention tasks was varied (Exp. III), again no effect of interpolated task difficulty was obtained, although decrements were observed relative to a rest control condition. When an unpaced interpolated task having no accuracy requirements was used (Exp. IV), decrements were again observed. These results fail to support a limited capacity channel model of short-term memory even when the model is limited to small movement distances. Results also yield only weak support for a trace interaction model as proposed by R. L. Pepper and L. M. Herman in 1970.

Interest in the effects of interpolated activity upon retention of a discrete kinesthetic movement has increased since Adams and Dijkstra (1966) offered a trace decay theory of short-term motor memory (STMM). Several experiments have attempted to demonstrate interference effects in STMM with varied results. Earlier studies were unable to demonstrate interference effects (Posner, 1967; Posner & Konick, 1966), but more recent studies, (Ascoli & Schmidt, 1969; Craft & Hinrichs, 1971; Herman & Bailey, 1970; Pepper & Herman, 1970; Stelmach & Wilson, 1970; Williams, Beaver, Spence, & Rundel, 1969) have shown decrements in recall when an interpolated motor or verbal task was used.

Broadbent's (1958) model of a limited capacity channel has played a central role in much of this research. A limited capacity channel model of the rehearsal process permits rehearsal only at the cost of available channel capacity. If the information to be retained is encoded so that less than full channel capacity is required for rehearsal, the limited channel model predicts no forgetting during the storage phase. A basic postulate of the limited capacity channel model is the additivity of information. If an interpolated task is required, the channel capacity used to perform the interpolated task is no longer available for processing information that is to be retained (memory information). Thus in order for memory information to be correctly maintained the sum of information per unit time required for both memory and interpolated tasks cannot exceed channel capacity. If this sum is greater than channel capacity, decrements will be observed in either the memory task, the interpolated task, or both. However, if *S*s are constrained to perform the interpolated task at a fixed information rate, the decrement will be observed only in the memory task if error rates are negligible or comparable for different levels of interpolated task difficulty.

Two studies by Posner (Posner, 1967; Posner & Konick, 1966) appeared to support a trace decay theory of STMM since

---

[1] This research was supported by XR and XL grants from the Purdue Research Foundation. The assistance of Edward Geiselman and Frederic Widlak in collecting and analyzing data is gratefully acknowledged. I am especially indebted to Louis Herman for having both the patience and fortitude to critically evaluate three previous versions of the article.

[2] Requests for reprints may be addressed to Barry H. Kantowitz, Department of Psychological Sciences, Purdue University, Lafayette, Indiana, 47907.

no effect of interpolated task upon recall was reported for kinesthetic information, although interference was demonstrated for visual information. The interpolated tasks were chosen to provide varying amounts of information reduction (Posner & Rossman, 1965). Digit pairs were read by *S at his own pace* and the appropriate transformation was written by *S* under each digit pair. However, such a self-paced interpolated task is not an adequate test of a limited channel model since the interpolated tasks are specified in terms of bits, whereas channel capacity is measured in bits per unit time (Broadbent, 1965). In a self-paced mode, *S*s typically perform fewer of the more difficult interpolated information reductions (Posner & Rossman, 1965). Hence, actual channel capacity (bits/sec) used in processing the different interpolated tasks in the two Posner studies may have been quite similar.

Williams et al. (1969) also used an unpaced interpolated task and found no effect of interpolated task difficulty upon recall. Their interpolated tasks were rest, recording, and addition. The recording task involved no information reduction, while the addition task supposedly involved an information reduction of 1.86 bits per pair of movements, based upon the number and probabilities of the set of pairs presented. This use of the information metric is incomplete since it fails to take into account the information generated by the motor system in performing the indicated movements (Fitts, 1954). Thus, it is not surprising that over the 30-sec. retention interval no differences were found between recording and addition interpolated tasks. However, this error was significantly greater than the error for the rest interpolated condition. Thus, it appears that interference effects in STMM can be demonstrated when the interpolated and motor tasks are extremely similar, as was the case in the Williams et al. study. However, a limited capacity model predicts interference effects that are related to task difficulty (in bits/sec). The Williams et al. experiment did not adequately test a task difficulty dimension since the interpolated tasks were unpaced and unrelated to information generated by the motor system.

Stelmach and Wilson (1970) used interpolated tasks of rest and addition similar to those of Williams et al., but their addition interpolated task was paced at a rate of three movements every 5 sec. Over a 20-sec. retention interval, absolute recall error was significantly higher for the paced addition task as compared to the rest condition. Furthermore, in accordance with findings of Posner and Keele (1969) interpolated activity had a greater effect on shorter retention movements. While use of a paced interpolated task is a great improvement over previous research, the locus of interference effects is still unresolved. Since the interpolated task was extremely similar to the retention task and was unrelated to information generated by the motor system, it is unclear that task difficulty rather than similarity accounts for interference effects as predicted by a limited capacity model. Furthermore, only one level of information reduction, other than a rest condition, was incorporated. Since performing no interpolated task is qualitatively different from performing some interpolated task, a more adequate experimental design should incorporate *at least two* levels of interpolated task difficulty in addition to a rest control condition. Stelmach and Bruce (1970) did use two paced interpolated tasks other than rest: addition and mimicking, which required *S* to duplicate pairs of angles. With absolute error as the dependent variable, no difference was obtained between addition and mimicking conditions, although both conditions revealed greater error than the rest condition. With algebraic error as the dependent variable, only the addition task differed from the rest condition, exhibiting greater overshoot error. However, it is unclear that comparisons between addition and mimicking conditions are not related to the additional movement required in the addition task rather than the (possibly) greater difficulty of the addition task; limitations of the addition task have already been noted.

Other studies which varied interpolated task similarity by having *S*s generate either a movement (Ascoli & Schmidt, 1969; Craft & Hinrichs, 1971) or a force (Herman & Bailey, 1970; Pepper & Herman, 1970) of a different extent than the memory information, but otherwise, qualitatively identical to the retention task, also have found interference effects related to task similarity, although the dimension of task difficulty remained untested in these studies.

However, while extreme similarity between interpolated and retention tasks may be a sufficient condition for STMM interference effects, it is not a necessary condition. Pepper and Herman (1970) and Herman and Bailey (1970) were able to demonstrate interference effects with a verbal interpolated task which required *S*s to count backwards by three's. Furthermore, Pepper and Herman reanalyzed data obtained by Posner (1967) and were able to show a pronounced effect of digit classification as compared with a rest control condition. Again the dimension of task difficulty was not tested. Both Herman studies used only one level of the verbal interpolated task, e.g., counting by three's, whereas at least two levels are needed to investigate task difficulty.

Thus, no study to date has adequately tested interpolated task difficulty as a source of interference in STMM. Such an omission is particularly astonishing when it is realized that Broadbent's (1958) limited capacity channel has often been cited in much of this research. The present experiments attempt to further localize interference effects in STMM by using two paced interpolated tasks which, although both motor tasks, differ in similarity to the retention task. One interpolated task requires tapping back and forth between two targets; the information generated by the motor system in performing this information conservation task is readily calculated and can be controlled by varying the size of the target plates and the distance between plates (Fitts, 1954). The similar interpolated task requires sliding, rather than tapping, between two targets; again information generated per unit of time is easily determined. In addition to a rest condition, two levels of difficulty are investigated in each case.

EXPERIMENT I

Method

Subjects.—The *S*s were 96 undergraduate students who participated to fulfill course requirements in introductory psychology. All *S*s were right-handed. Half of the *S*s were males and half females. Each *S* experienced all experimental conditions.

Apparatus.—The device used was similar to that of Adams and Dijkstra (1966). A slide and stop were mounted on a concealed rod. Following onset of a signal light, *S* moved the slide until his progress was arrested by the stop, thus defining the distance to be retained. The slide was hydraulically returned to its starting position where it was released by *S*; the gas-operated return was used to ensure equal return times for all conditions. A large "bib" was placed around *S*'s neck to prevent any visual cues, such as arm position, from being used.

The interpolated tapping task was paced by a metronome at a rate of two taps per second. Two pairs of targets were used. The target pair used in the Low information rate interpolated task consisted of two 2-in. (5.08-cm.)-width plates separated by an edge-to-edge distance of ¼ in. (.64 cm.); this combination yielded an index of difficulty of 1.17 bits (Fitts, 1954). The target pair used in the High information rate interpolated task consisted of two ½-in. (1.27-cm.) plates separated by an edge-to-edge distance of 7½ in. (19.05 cm.), yielding an index of difficulty of 5 bits. Since movements were performed at a 2/sec rate, these target pairs yielded information transmission rates of 2.3 bits/sec (Low) and 10 bits/sec (High) if performance was without error. Tapping errors (e.g., undershoots and overshoots) were electrically recorded.

Procedure.—A constant retention interval of 20 sec. was used in all conditions. One major variable was activity during this interval. Four combinations of tapping (T) and resting (R) were used: RR indicating 20 sec. of rest; TR indicating 10 sec. of tapping followed by 10 sec. of rest; RT indicating 10 sec. of rest followed by 10 sec. of tapping; TT indicating 20 sec. of tapping. All 24 orders (4!) of these four conditions were replicated four times between *S*s. The other task variable was information rate of the interpolated tapping task which was either high (H) or low (L). The *S*s were instructed to minimize tapping errors and were given tapping practice until their error rate was below 2%; approximately 4 min. of practice was sufficient to attain this criterion. Information rate was partially counterbalanced with half of the *S*s starting with the H task and half with the L task. Each block consisted of 10 trials with movement amplitudes (i.e., retention distances) of from 15 to 24 cm. in 1-cm. steps; amplitude was sampled randomly without replacement for each block. Thus, each *S* experienced eight 10-trial blocks.

Results

Both algebraic and absolute mean recall errors are given in Table 1. For algebraic errors, a positive sign indicates an overshoot error and a negative sign an undershoot error. Analyses of variance for algebraic error revealed highly significant effects of movement amplitude, $F\ (9, 792) = 16.66$, $p < .001$; the two shortest movement amplitudes revealed undershooting which changed to progressively greater overshooting as amplitude increased. A highly significant Interpolated Task Difficulty × Order interaction, $F\ (1, 88) = 14.31$, $p < .001$, revealed that *S*s who experienced the L tapping task followed by the H task had a greater slope than *S*s who received the opposite order when error was plotted as a function of interpolated task ordinal position.

Since such an interaction with order often indicates an undesirable asymmetric transfer effect (Poulton & Freeman, 1966), an additional analysis of algebraic error was performed for only the first half of the data, thus making interpolated task difficulty a between-*S*, rather than a within-*S*, variable. Effects of amplitude were again highly significant, $F\ (9, 792) = 4.79$, $p < .001$; effects of activity were also significant, $F\ (3, 264) = 3.07$, $p < .05$. Comparisons between Cond. RR and TR, $t\ (264) = 1.96$, and RR and TT, $t\ (264) = 1.62$, both *p*s > .05, approached significance; comparisons between Cond. TR and RT, $t\ (264) = 2.57$, $p < .02$, and RT versus TT, $t\ (264) = 2.23$, $p < .05$, were significant. Effects of interpolated task difficulty, although showing greater overshoot for the H interpolated task (+.284 cm. vs. + .043 cm.), were not significant, $F\ (1, 88) = .97$. Similar analysis of absolute recall error found no effects which were significant at the .05 level for either first half only or both halves combined.

TABLE 1

MEAN RECALL ERROR FOR KINESTHETIC MODALITY

Interpolated task difficulty	Activity			
	RR	RT	TR	TT
	Both halves combined			
Low: 2.3 bits/sec	2.04 +.346	2.07 +.086	2.02 +.516	1.88 +.225
High: 10 bits/sec	1.92 +.295	1.91 +.426	1.83 +.374	1.95 +.312
	First half only			
Low: 2.3 bits/sec	1.96 −.089	2.09 −.165	2.01 +.274	2.05 +.152
High: 10 bits/sec	1.90 +.222	2.01 +.139	1.79 +.370	1.97 +.404
$\bar{X}$[a]	1.93 +.067	2.05 −.013	1.90 +.322	2.01 +.278

Note.—Error is measured in centimeters.
[a] Mean is given for first-half-only data.

Thus, Exp. I fails to demonstrate any decrement in retention due to interpolated task difficulty. Error rates in both interpolated tasks were below 2% so that this finding cannot be attributed to less information generated in the H task. Both tapping tasks were paced at the same rate so that transmitted information was greater for the H task since *S*s could not slow down their rate of tapping as in a self-paced task. Interpolated tapping did cause increased overshooting when tapping was performed for the first 10 sec. of the delay interval contrasted with a slight undershooting when tapping was performed for the last 10 sec. of the 20-sec. delay interval.

EXPERIMENT II

Experiment II investigates recall error when visual cues are available in addition to the kinesthetic cues used in Exp. I. Posner (1967), using a lever-positioning task, found interference effects in visual STMM but not in kinesthetic STMM. This is an important finding since it suggests that visual information can be maintained by rehearsal, while kinesthetic information cannot. Although this finding might also suggest visual STMM to be more interesting than kinesthetic STMM, later studies have focused exclusively upon kinesthetic STMM so that the amount of central processing capacity required to rehearse visual coding has yet to be established.

On the basis of two studies (Posner, 1967; Posner & Konick, 1966), Posner agrees with Adams and Dijkstra (1966) that kinesthetic information is coded nonverbally. Posner also maintains that such is the case for visual information as well. However, since visual information led to improvement over the unfilled retention interval for the visual location condition (Posner, 1967), a verbal code cannot be rejected. Posner's conclusion regarding the storage code for visual information is thus based upon the introspective reports of his *S*s.

Better estimates of the amount of central processing capacity needed for visual rehearsal and of the nature of a visual storage code can be achieved by using two levels of interpolated task difficulty in addition to rest. The present experiment improves upon Posner's earlier (1967) methodology by using two paced interpolated tasks in place of Posner's unpaced digit classification task; furthermore, the present study also scored performance on the interpolated tasks to ensure that the tasks were performed accurately.

Method

Method and procedure were similar to that in Exp. I, except that the cover and bib which prevented seeing the rod and slider were removed. The cover was replaced by a cheese-cloth screen which permitted *S* to see both his hand and the rod. This screen was used because preliminary pilot work (Kantowitz, 1969) indicated that the small scratch marks caused by moving the slide were potent visual cues which were obviated by inclusion of the screen. The *S*s were 72 male undergraduates from the same population as those in Exp. I. Other procedural details were identical to those of Exp I.

Results

Both algebraic and absolute recall errors are given in Table 2. For algebraic error effects of amplitude, F (9, 630) = 3.02, $p < .01$, and order (LH vs. HL), F (1, 70) = 6.22, $p < .05$, were significant; the interaction of order and interpolated task difficulty was not significant, F (1, 70) = 1.11, $p > .05$. No other main effects or interactions reached the .05 level of significance. Effects of amplitude revealed a very small undershoot (− .002 cm.) for the smallest amplitude after which progressively greater overshoots occurred as amplitude increased. Order LH exhibited greater overshoot (+ .414 cm.) than order HL (+ .135 cm.).

TABLE 2

MEAN RECALL ERROR FOR VISUAL + KINESTHETIC MODALITY

Interpolated task difficulty	Activity			
	RR	RT	TR	TT
Low: 2.3 bits/sec	.825 +.276	.935 +.162	.999 +.305	1.056 +.189
High: 10 bits/sec	.821 +.297	1.025 +.334	.978 +.409	.924 +.223
$\bar{X}$	.823 +.286	.980 +.248	.988 +.357	.990 +.206

Note.—Error is measured in centimeters.

For absolute error, only effects of activity, F (3, 210) = 5.32, and amplitude, F (9, 630) = 3.38, both $p\text{s} < .01$, were significant. Effects of order, F (1, 70) = 2.29, and its interaction with interpolated task difficulty, F (1, 70) < 1.0, failed to reach the .05 level of significance. Error was a shallow U-shaped function of amplitude with greatest error at the four largest amplitudes. Condition RR had significantly smaller error than any of the three remaining activity conditions, $t\text{s}$ (210) ≥ 3.14, $p < .01$; Cond. RT, TR, and TT did not differ, $t\text{s}$ (210) < 1.0.

As was the case with Exp. I, no evidence for recall decrements caused by difficulty of the interpolated task was found. With absolute error as the dependent variable, a rest condition showed smaller error than any tapping condition. While these results do replicate Posner's (1967) findings since an interpolated task increased absolute error, they do not support his conclusion that rehearsal of visual information is dependent upon available central processing capacity; if such were the case, effects of interpolated task difficulty should have been detected.

EXPERIMENT III

The preceding experiments have found no evidence for STMM interference which is a function of interpolated task difficulty for either kinesthetic or visual + kinesthetic modalities. However, while an interpolated motor task was used, it is possible to use an additional interpolated motor task which is even more similar to the retention task. If interference effects are due to more than the simple interpolation of some intervening activity, regardless of the difficulty of this activity or its similarity to the retention task, at least two hypotheses may be formulated: (*a*) Increased error should be associated with the more similar interpolated task if similarity governs interference in STMM. Thus, a significant main effect of similarity is predicted. (*b*) In order to retain a limited capacity channel model, some effect of task difficulty need be demonstrated. In light of the previous experiments, it may be that task difficulty is effective only for highly similar interpolated tasks. This hypothesis predicts an interaction between task difficulty and task similarity.

Method

Subjects.—The *S*s were 96 undergraduate students who participated to fulfill course requirements in introductory psychology. All *S*s were right-handed males.

Apparatus.—A slide to which a pointer was affixed was mounted on a concealed horizontal rod. Targets placed under the pointer controlled information generated as the slide was moved back and forth between the two targets. Overshoot errors were electrically recorded and undershoot errors were observed by *E*. Targets were matched in width and spacing to targets used in the tapping task of previous experiments. This new apparatus was used for the similar interpolated task conditions. The remainder of the apparatus was that used in the previous experiments.

Procedure.—One group of 48 *S*s was assigned to the kinesthetic modality and a bib was used as in Exp. I. The two remaining groups of 24 *S*s each were assigned to either the visual modality as in Exp. II or to a kinesthetic condition with no bib being used to minimize visual cues (KNB). A constant retention interval of 20 sec. was used as in Exp. I and II. However, this interval was not bifurcated. All *S*s underwent all factorial combinations of two levels of interpolated task similarity (tapping and sliding) with two levels of interpolated task difficulty (Low and High as in previous experiments). All 24 orders of the four conditions were used with 2 *S*s within each order for the kinesthetic condition. In addition, a 20-sec. rest condition occurred as the middle block of trials for all *S*s; contrary to the previous experiments, *S*s were not allowed to disengage by removing their hand from the slide during the rest period. Other details of procedure were identical with those of the previous experiments.

Results

Both algebraic and absolute mean recall errors are given in Table 3. Separate analyses of variance for the three groups failed to reveal significant effects at the .05 level for task difficulty, task similarity, or blocks, or their interactions for either dependent variable when the rest control condition was omitted from the analysis. Hence, neither of the two hypotheses formulated above was supported. When three additional analyses of variance were conducted with the rest condition included, effects of interpolated activity

TABLE 3

MEAN ABSOLUTE AND ALGEBRAIC RECALL ERROR

Group	Interpolated activity				
	Rest	Tapping		Sliding	
		Low	High	Low	High
Kinesthetic	1.75 (1.47) +.29 (2.27)	2.34 (1.79) +.07 (2.95)	2.28 (1.83) +.17 (2.92)	2.16 (1.76) −.08 (2.79)	2.27 (1.78) +.08 (2.88)
Visual + Kinesthetic	.62 (.61) +.32 (.81)	1.12 (.90) +.22 (1.42)	1.31 (1.30) +.32 (1.82)	1.17 (.91) +.46 (1.41)	1.12 (1.04) +.49 (1.45)
KNB	1.20 (1.10) +.14 (1.62)	1.77 (1.54) −.09 (2.34)	1.89 (1.46) +.17 (2.38)	1.83 (1.57) −.12 (2.41)	1.94 (1.56) +.37 (2.46)

Note.—Error is measured in centimeters. Numbers in parentheses are standard deviations. $N = 48$ for the Kinesthetic group; $N = 24$ for both the Visual + Kinesthetic and the KNB groups.

were significant for absolute error for kinesthetic, F (4, 96) = 5.48, $p < .001$, visual, F (4, 92) = 7.74, $p < .001$, and KNB, F (4, 92) = 4.92, $p < .01$, conditions; the effect of blocks was significant only for the kinesthetic condition, F (4, 96) = 5.48, $p < .001$. Three similar analyses of variance for algebraic error found no main effect other than blocks, F (4, 96) = 4.76, $p < .01$, to reach the .05 level of significance. Thus, Exp. III replicates and extends the findings of the two preceding experiments in that interference effects were unrelated to interpolated task difficulty. The improved recall in the present rest condition can be attributed to the requirement that *S*s do not disengage the slide during the retention interval.

Of interest, however, is an unexpected significant Task Difficulty × Movement Amplitude interaction, F (9, 216) = 3.05, $p < .01$. Except at the shortest movement amplitude, the two interpolated task difficulties exhibit a crude mirror symmetry when plotted, with each overshoot error being balanced by a slightly smaller undershoot error. While the finding of greater overshoot error for the more difficult interpolated task and shortest movement amplitudes is in agreement with Posner and Keele (1969), an even greater difference is found for distances of 20 and 21 cm., and greater overshoot is associated with the less difficult interpolated task for three movement amplitudes. Thus, on balance these data do not support a limited channel model which is valid for only short retention movements.

EXPERIMENT IV

The preceding experiments have failed to demonstrate interference effects which are related to the difficulty or qualitative similarity of the interpolated task. However, effects of interpolated activity were found since retention generally proved superior for a rest condition during which no interpolated activity need be performed. The present experiment attempts to demonstrate that even the simplest interpolated activity, e.g., one that has neither accuracy nor pacing requirements, will cause interference effects relative to a rest control condition. The interpolated task chosen is roughly comparable to interpolated tasks used by Williams et al. (1969) since it involves repeated movements of the slide; however, the present study does not claim to vary information.

Method

Twenty-four male undergraduates were assigned to each of three conditions. In the Kinesthetic (K) condition the apparatus was arranged as in Exp. I. In the two Visual + Kinesthetic (V) conditions, it was arranged as in Exp. II. In the standard V condition, *S*s always faced the slide and rod during the 20-sec. delay interval, even when performing the interpolated task. In the Visual Look-Away (VLA) condition, *S*s did not view the slide and rod during the delay interval.

The interpolated task involved moving the slide back and forth in some haphazard manner. The *S*s were free to wave the slide about in any way they chose. The same four activity conditions were again used: RR, RT, TR, and TT, where T now indicates haphazard movement of the slide rather than paced tapping as before. Other procedural details were identical to those of Exp. I and II; e.g., *S*s were allowed to disengage from the slide.

Results

Mean absolute and algebraic errors for the three conditions are given in Table 4. For the K group with absolute error as the dependent variable, effects of activity were significant, F (3, 69) = 4.61, $p < .01$. Conditions RR and RT did not differ significantly, t (69) = 1.76, $p > .05$, but RR did differ significantly from TR, t (69) = 2.21, $p < .05$, and TT, t (69) = 3.69, $p < .001$. No other differences reached the .05 level of significance. With algebraic error as the dependent variable, effects of

TABLE 4

ACTIVITY EFFECTS ON MEAN RECALL ERROR

Group	Activity			
	RR	RT	TR	TT
Kinesthetic	1.50 −.526	1.89 −.041	1.99 +.073	2.32 +.712
Visual + Kinesthetic	.904 +.463	1.09 +.477	1.06 +.422	.862 +.252
VLA	1.01 +.436	1.28 +.455	1.60 +.134	1.58 +.383

Note:—Error is measured in centimeters.

activity were again significant, F (3, 69) = 4.59, $p < .01$. Only Cond. RR and TT, t (69) = 3.68, $p < .001$, and RT and TT, t (69) = 2.24, $p < .05$, differed significantly at the .05 level.

For the V group with absolute error as the dependent variable effects of both activity, F (3, 69) = 2.82, $p < .05$, and amplitude, F (9, 207) = 2.10, $p < .05$, were significant. Condition RR did not differ significantly from any of the three remaining conditions, ts (69) $\leq$ 1.99, $p > .05$, but Cond. TT did differ significantly from Cond. TR, t (69) = 2.05, $p < .05$, and RT, t (69) = 2.44, $p < .02$. Since *S*s in this condition always viewed the slide and rod, the decreased error may indicate an overt rehearsal process maintained by visual feedback. Effects of amplitude revealed slightly greater error for shorter amplitudes. With algebraic error as the dependent variable, only effects of amplitude, F (9, 207) = 3.01, $p < .01$, were significant. These revealed moderate overshoot at the smallest amplitude (+ .33 cm), increased overshoot at the next three largest amplitudes, followed by decreasing overshoot as amplitude increased.

For the VLA group with absolute error as the dependent variable, effects of activity, F (3, 69) = 8.21, $p < .001$, and amplitude, F (9, 207) = 2.34, $p < .05$, were significant. Condition RR did not differ significantly from Cond. RT, t (69) = 1.98, $p > .05$, but did differ from Cond. TR, t (69) = 4.30, and TT, t (69) = 4.12, both ps $< .001$. Condition RT also differed from Cond. TR, t (69) = 2.32, $p < .05$, and TT, t (69) = 2.14, $p < .05$. Amplitude effects revealed greatest error for the three shortest amplitudes. With algebraic error as the dependent variable, only the Activity × Amplitude interaction, F (27, 621) = 1.87, $p < .01$, was significant. Condition RR was characterized by a large overshoot (+ .21 cm.) at the two greatest amplitudes; this was also true for Cond. RT. Conditions TR and TT, however, exhibited either undershoots or very small overshoots at the two greatest amplitudes.

Thus, the present experiment clearly demonstrates interference effects in both visual and kinesthetic modalities which are related only to the activity inherent in the interpolated task.

GENERAL DISCUSSION

The present experiments demonstrate that production of a motor response can be influenced by a variety of motor interpolated tasks. This finding holds for both paced and unpaced tasks. Furthermore, these effects are unrelated to the qualitative similarity of the interpolated motor task relative to the production task. Finally, the amount of information transmitted per unit time in a conservation task encompassing a range of 2.3 to 10 bits/sec does not affect retention of a motor item.

The present experiments offer no support for a limited capacity channel model of STMM. For both kinesthetic and visual + kinesthetic modalities difficulty of the interpolated task did not affect either absolute or algebraic recall error. Recent experiments (Posner & Keele, 1969; Stelmach & Wilson, 1970) have proposed a modified channel capacity model whereby only short movement amplitudes required channel capacity. The range of amplitudes used in the present studies qualify as short to medium when compared with prior studies (e.g., Adams & Dijkstra, 1966; Williams et al., 1969). Nevertheless, no interaction between amplitude and interpolated task difficulty was found for either modality for Exp. I and II. Inspection of the data of Exp. I and II revealed no evidence for any effect of interpolated task difficulty at even the shortest movement amplitudes. The interaction found in Exp. III was too complex to be consistent with this modified channel capacity model. Since both the Posner and Keele and Stelmach and Wilson studies used only one level of interpolated task, other than rest, their conclusions seem premature at best. Three possible explanations might be used to resolve this discrepancy. The present study used a linear positioning task as compared to the lever positioning task used by Posner and Keele as well as by Stelmach and Wilson. However, since effects of delay are comparable between the two kinds of positioning task (Kantowitz, 1969) and since both are discrete positioning tasks, it seems unlikely that this procedural difference can account for the two differing sets of results.

A second alternative is the difference between interpolated tasks: information reduction versus conservation. However, until at least two levels of information reduction are incorporated into the same experiment, it is logically impossible to evaluate possible effects of information reduction which are indeed related to the difficulty per se of the interpolated task, rather than to the qualitative difference between no information reduction and some information reduction. Schmidt and Ascoli (1970) have compared effects of a conversation task (counting forward by one's) and a reduction task (counting backwards by three's) in a proactive inhibition paradigm and found significantly greater absolute error for the reduction task. However, it would be dangerous to extrapolate this finding to retroactive inhibition paradigms, especially since Herman and Bailey (1970) found different proactive and retroactive effects in STMM.

A third explanation, suggested by the reviewer, concerns the attention demands of the interpolated tapping task. If the movement is automated, i.e., can be performed without requiring any attention, then no decrement would be expected in the memory task (Posner & Keele, 1969). This criticism merits discussion and can be approached on both logical (operational) and empirical grounds. By definition, the limited channel model demands an upper bound upon rate of transmitted information (Broadbent, 1965). Since the interpolated task generated rates of information flow it satisfies, a priori, the requirements necessary for test of a limited capacity model. An explanation of the present "all-or-none" findings based upon defining the interpolated task as highly automated renders the limited channel model exempt from this test only at the cost of limiting the generality of the model. The interpolated tapping task generates information and requires time. Rejecting the interpolated task implies that it is no longer sufficient to specify the information transmission rates of an interpolated task in terms of observable and measurable bits/sec since these (externally) determined information rates may not be monotonically related to the (unobservable) information rate within the limited channel itself. If the psychological (i.e., limited channel) demands of the interpolated task cannot be monotonically related to the measurable external task demands, it becomes exceedingly difficult to test the limited channel model; this renders the model less attractive. On empirical grounds, this investigator is unaware of any study which directly tests the hypothesis that the Fitts Law tapping task is highly automated. Most studies of dual-task performance have found a considerable degree of practice necessary before a task can be sufficiently automated to be combined with a second task without degrading performance (e.g., Herman, 1965). The present study used roughly 4 min. of tapping practice prior to the experiment proper. Nevertheless, this possibility cannot be rejected without empirical efforts, and further research utilizing the tapping task within a dual-task set is currently in progress.[3]

Since the limited channel model cannot explain the present findings, it is appropriate to examine alternative models. One recent model has been proposed by Pepper and Herman (1970). Only that portion of their model dealing with effects of interpolated activity will be discussed. Pepper and Herman suggested two major mechanisms to account for effects of interpolated task in STMM. When the interpolated task is qualitatively identical to the retention task (e.g., Craft and Hinrichs, 1971; Pepper and Herman 1970), the resultant STMM trace is a weighted mean of interpolated and original traces. However, for interpolated tasks which are not motor tasks related to the retention task (e.g., counting backwards), a muscle tension mechanism comes into prominence. The Pepper-Herman model is not sufficiently articulated to allow determining the relative effects of these two mechanisms in the present situation. Nevertheless, predictions can be made for each of the two mechanisms considered separately. The first mechanism would be sensitive to the extent of movement during the interpolated task; furthermore, effects of movement extent should be more pronounced in the sliding, rather than the tapping interpolated task since the sliding task is qualitatively similar to the retention task. Movement extent for the H interpolated task was approximately 19 cm., while movement extent for the L interpolated task was approximately 5 cm. Since retention extents varied over a range of 15 to 24 cm., the greater discrepancy in length between retention and L movements versus retention and H movements leads to a prediction of greater undershoot for the L interpolated task. Such an outcome was obtained only for Group KNB of Exp. III, but the finding was not statistically reliable. Perhaps however, the muscle tension mechanism is dominant so that trace interaction is relatively weak component in final response production. The muscle tension mechanism acts to augment the trace of a motor item as muscle tension is increased. It is reasonable to associate a greater level of general muscle tension with the more difficult H interpolated task rather than with the L interpolated task; hence, greater overshoot would be predicted for the H task. While some evidence for this hypothesis was obtained (Tables 1, 2, and 3), it was not statistically reliable. The lack of significant effects of task difficulty, despite the large number of observations, argues against the muscle tension hypothesis unless it is further assumed that muscle tension rises quite rapidly to asymptote as interpolated activity is increased. If so, the explanatory value of the hypothetical muscle tension mechanism is diminished as the reciprocal of the muscle tension activity function, since rapid attainment of asymptotic muscle tension precludes a graded effect of the mechanism; i.e., muscle tension then operates differentially only over a small range of activity so that an "all-or-none" trace augmentation is operative. Thus, the inability of either mechanism to explain present data limits the generality of the Pepper-Herman model.

---

[3] This research has been completed and shows that the tapping task cannot be combined with another task without a decrement in performance on the other task. Thus the tapping task does demand attention.

The Pepper-Herman (1970) model also has some difficulty in accounting for effects of interresponse interval (IRI) in retroactive inhibition. Herman and Bailey (1970) found a 14-sec. IRI to produce more interference than a 4-sec. IRI. This was explained for motor interpolated tasks via the first (trace interaction) mechanism with the assumption that the weight of a motor act is proportional to its distance in time from the retention test. The second (muscle tension) mechanism is subject to trace decay effects in that the residual level of muscle tension aroused by interpolated activity is assumed to decay over time. Thus, for longer IRIs when the interpolated task is temporally closer to the retention test, the level of muscle tension should be higher than for short IRIs, where muscle tension has a longer period before the retention test and can thus decay to a lower level. The present experiments varied IRI by dichotomizing the retention interval; i.e., conditions which called for immediate initiation of the interpolated task (e.g., TR, TT) utilized an IRI of 0 sec., while Cond. RT used an IRI of 10 sec. Hence, Herman's model (Herman & Bailey, 1970) predicts greater overshoot for Cond. RT relative to Cond. TR. Significantly greater overshoot was obtained (Exp. I) for Cond. TR relative to Cond. RT, an outcome which contradicts the data and explanation given by Herman and Bailey. However, no effect of IRI was obtained when both visual and kinesthetic information was used (Exp. II). When an unpaced motor interpolated task was used (Exp. IV), the IRI variable influenced only condition VLA with algebraic error as the dependent variable. Again, the absence of an IRI effect in the Kinesthetic condition contradicts findings of Herman and Bailey.[4]

Given the differences in results when the present studies are compared to the earlier studies conducted by Herman (Herman & Bailey, 1970; Pepper & Herman, 1970), additional theorizing seems hazardous at this point. It is, however, clear that the qualitative relation between retention and interpolated tasks in STMM needs further investigation. A trace interaction model may hold only for interpolated tasks which are subsets of the retention task. For the kinds of interpolated motor tasks investigated in the present experiments, neither a trace interaction nor a limited capacity model appear viable; similarly a muscle tension explanation as suggested by Pepper and Herman may not be able to explain IRI effects without additional emendation.

---

[4] In a personal communication (January 1972), L. M. Herman has suggested a possibility in defense of his theory. In Herman's work, the end of the criterion force was always followed in the short IRI condition by a 3–4 sec. unfilled interval before the initiation of the criterion force. The present experiments required immediate initiation of tapping in Cond. TR. Thus, the immediate onset of the interpolated task in the present experiments may have interferred with consolidation of the proprioceptive trace associated with the initial memory task.

REFERENCES

Adams, J. A. & Dijkstra, S. Short-term memory, for motor responses. *Journal of Experimental Psychology*, 1966, **71**, 314–318.

Ascoli, K. M., & Schmidt, R. A. Proactive interference in short-term motor retention. *Journal of Motor Behavior*, 1969, 1, 29–36.

Broadbent, D. E. *Perception and communication.* London: Pergamon Press, 1958.

Broadbent, D. E. Applications of information theory and decision theory to human perception and reaction. In N. Wiener & J. P. Schade (Eds.), *Cybernetics of the nervous system.* Amsterdam: Elsevier, 1965.

Craft, J. L., & Hinrichs, J. V. Short-term retention of simple motor responses: Similarity of prior and succeeding responses. *Journal of Experimental Psychology*, 1971, **87**, 297–302.

Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 1954, **47**, 381–391.

Herman, L. M. Study of the single channel hypothesis and input regulation within a continuous, simultaneous task situation. *Quarterly Journal of Experimental Psychology*, 1965, **17**, 37–46.

Herman, L. M., & Bailey, D. R. Comparative effects of retroactive and proactive interference in motor short-term memory. *Journal of Experimental Psychology*, 1970, **86**, 407–415.

Kantowitz, B. H. Recognition and recall in short-term motor memory. Unpublished doctoral dissertation, University of Wisconsin, 1969,

Pepper, R. L., & Herman, L. M. Decay and interference effects in the short-term retention of a discrete motor act. *Journal of Experimental Psychology*, 1970, **83** (2, Pt. 2).

Posner, M. I. Characteristics of visual and kinesthetic memory codes. *Journal of Experimental Psychology*, 1967, **75**, 103–107.

Posner, M. I., & Keele, S. W. Attention demands of movements. *Proceedings of the XVIth International Congress of Applied Psychology*. Amsterdam: Swets & Zeittinger, 1969.

Posner, M. I., & Konick, A. F. Short-term retention of visual and kinesthetic information. *Organizational Behavior and Human Performance*, 1966, **1**, 71–88.

Posner, M. I., & Rossman, E. The effect of size and location of interpolated information reducing transformations upon short-term retention. *Journal of Experimental Psychology*, 1965, **70**, 491–505.

Poulton, E. C., & Freeman, P. R. Unwanted asymmetrical transfer effects with balanced experimental designs. *Psychological Bulletin*, 1966, **66**, 1–8.

Schmidt, R. E., & Ascoli, K. M. Attention demand during storage of traces in motor short-term memory. *Acta Psychologica*, 1970, **34**, 497–504.

Stelmach, G. E., & Bruce, J. R. Recall load in STM. *Psychonomic Science*, 1970, **21**, 205–207.

Stelmach, G. E., & Wilson, M. Kinesthetic retention, movement extent, and information processing. *Journal of Experimental Psychology*, 1970, **85**, 425–430.

Williams, H. L., Beaver, W. S., Spence, M. T., & Rundell, O. R. Digital and kinesthetic memory with interpolated information processing. *Journal of Experimental Psychology*, 1969, **80**, 530–536.

(Received November 22, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 275–284

SENTENCE–PICTURE COMPARISON:
A TEST OF ADDITIVITY OF PROCESSING TIME FOR FEATURE MATCHING AND NEGATION CODING[1]

LESTER E. KRUEGER[2]

City College of the City University of New York

According to T. Trabasso, H. Rollins, and E. Shaughnessy and to W. G. Chase and H. H. Clark, when *S* decides whether a sentence accurately describes a feature of a picture, he performs two independent, serial operations, of which contributions to total processing time are additive. The *S* checks (*a*) whether the sentence representation matches the picture representation on the feature and (*b*) whether a negative is present in the sentence. In the present study, *S* compared a sentence (e.g., IS NORTH, ISN'T NORTH, IS EAST) with the location of a circle. Contrary to the serial model, (*a*) mean response time (RT) for true-negative sentences was shorter than was predicted based on RT for false affirmatives and false negatives, and (*b*) RT was longer if the features of the sentence and circle matched on axis (horizontal, vertical) and differed as to specific direction than if they differed on both axis and specific direction, and this axis-match effect was larger if the sentence was negative. The results support a model in which feature matching and negation coding are parallel operations sharing a limited-capacity central processor.

The sentence–picture comparison task requires *S* to decide, as quickly as possible, whether a sentence correctly describes a feature of a picture. Since the sentence typically can be either affirmative or negative, *S* must consider both the value of the relevant feature in the sentence and whether the value has been negated. For example, a picture of a red ball is accurately described by the sentence, "The ball is red," and the sentence, "The ball is not green." In the first case there is a feature match between sentence and picture. In the second case there is a feature mismatch that is canceled by a negative in the sentence.

Recently, Trabasso, Rollins, & Shaughnessy (1971) and Chase and Clark (in press) have independently proposed an additive model of sentence–picture comparison, according to which the picture and sentence are each encoded separately, and then their representations are compared piece by piece, with each operation at the comparison stage adding a separate increment to total processing time. The picture is assumed to be encoded as an affirmative proposition and the sentence, when negative, to be represented by the feature plus a negative indicator. At the comparison stage, there are two independent, serial operations: (*a*) feature matching and (*b*) cancellation of the negative, which is referred to as negation coding. A running tally of the two operations is kept by *S*'s response index, which can take the value "true" or "false," and which is initially set at true. At the feature-matching substage, a feature mismatch between picture and sentence takes longer to process than a feature match, according to the additive model, because a mismatch, but not a match, requires the response index to be reset.[3] Similarly, at the negation-coding substage, a negative sentence takes longer to process than an affirmative sentence because a negative requires the response index to be reset.[4] In some cases the negative may be removed by a transformation of the sentence before the comparison stage.

---

[1] This study was supported by National Institutes of Health Grant HD04869-02. The author is grateful to Paul Kadis and Sheri Berenbaum for collecting and tabulating the data, to Phillip Liss and Ira T. Kaplan for many very helpful discussions, and to David E. Meyer for his very helpful criticisms and comments on an earlier draft of this paper.

[2] Requests for reprints should be sent to Lester E. Krueger, Department of Psychology, City College of the City University of New York, New York, New York 10031.

[3] A feature mismatch may cause *S* to double-check and do other things in addition to resetting the response index. For simplicity of exposition, the time required for these other processes will be lumped under that for resetting the response index. Further, feature matching may require more than one resetting of the response index when the sentence describes two objects, such as, "plus above star" (cf. Chase & Clark, in press). The present paper, however, is concerned only with sentences that describe a single object.

When the picture is presented before the sentence, according to Trabasso et al. (1971) and to Chase and Clark (in press), the negative is preserved in the representation of the sentence. In this case, (*a*) a false-affirmative sentence (e.g., "The ball is green," with pictured ball being red) takes longer than a true-affirmative sentence ("The ball is red") because it involves a feature mismatch and thus a resetting of the response index, (*b*) a false negative ("The ball is not red") takes longer than the true affirmative because the presence of the negative also causes *S* to reset his response index, and (*c*) a true negative takes longest of all because the response index must be reset twice, once because of a feature mismatch and once because of a negative.

The crucial finding for the additive model of Trabasso et al. (1971) and of Chase and Clark (in press) is that the mean response time (RT) for the true negative has been found to be well predicted by the sum of (*a*) the true-affirmative RT, (*b*) the feature-mismatch increment (false-affirmative RT–true-affirmative RT), and (*c*) the negation increment (false-negative RT–true-affirmative RT). The additivity and lack of interaction for processing time for feature matching and negation coding suggest that there are two independent processing stages (cf. Sternberg, 1969).

Under certain other, less typical circumstances (i.e., the sentence is shown before the picture and the feature values are binary; it also may be necessary for *S*s to be well practiced, to be adults, and to be specially instructed), *S* may remove the negative from the sentence representation before the comparison stage. If the two feature values are red and green, for example, then the sentence, "The ball is not red," may be transformed and stored as, "The ball is green." In this case, (*a*) a false-affirmative sentence takes longer to verify than a true-affirmative sentence because of the feature mismatch; (*b*) a true negative takes longer to verify than a true affirmative because of the time taken to transform the sentence in order to remove the negative; and (*c*) a false negative takes longest of all because it involves both a resetting of the response index, due to a feature mismatch, and a sentence transformation to remove the negative. Thus, whether true negatives are verified faster or slower than false negatives depends on whether the negation is handled by a sentence transformation or by a resetting of the response index.

Some evidence indicates, however, that negation is not handled by a separate operation. Meyer (in press) found evidence against the additive model as applied to comparisons between sentences and information in memory. He had *S*s judge the truthfulness of logical propositions concerning familiar semantic categories, e.g., "Some doctors are males," and "No elms are metals." His results showed that (*a*) category size and the set relation between categories influenced the duration of memory retrieval, as reflected by RT, and (*b*) the effects of these variables were greater for negative than affirmative sentences. Thus, difficulty of memory retrieval interacted with sentence form (affirmative vs. negative).

EXPERIMENTS I–IV

The present study tests whether feature matching and negation coding involve separate, additive operations. A sentence specifying direction (e.g., IS NORTH, IS EAST, ISN'T NORTH) was presented with a circle that was either above (i.e., north), below (south), left (west), or right (east) of the sentence. Although the sentence and circle sometimes were presented simultaneously in the present experiments, rather than having the picture shown first, and *S*s were not told what to look at first, the fact that the direction feature was nonbinary, taking one of four values rather than two, meant that *S*s could not readily transform the sentence to remove the negative. Thus, according to the additive model, a false affirmative, which involves a feature mismatch, and a false negative, which involves a negative, ought to be verified more slowly than a true affirmative because each involves one resetting of the response index. The RT for a true negative ought to be longest of all because the response index is reset twice, since it involves both a feature mismatch and a negative. Thus, there ought to be an Affirmative–Negative × True–False interaction for RT, and, in addition, the RT for true negatives should exceed the base-line RT for true affirmatives by an amount equal to the sum of the increments for false affirmatives and false negatives over true affirmatives.

---

[4] Even if the negative is not removed by a sentence transformation, a negative sentence still may take longer to process at the encoding stage, as well as at the comparison stage. For simplicity of exposition, the extra time required for encoding is lumped under that for resetting the response index. The key notion of the additive model does not concern where negation is handled, but that negation adds a separate component to total processing time.

The present study also provides a second test of the additive model. False affirmatives (e.g., IS NORTH, with circle at the west) and true negatives (e.g., ISN'T NORTH, with circle at the west) both involve a feature mismatch and differ in that only the true negative involves negation. Some feature mismatches were more difficult than others because they involved an axis match: the sentence and circle differed on direction but agreed on horizontal or vertical axis, e.g., IS NORTH, with circle to south. If feature matching and negation coding are truly separate additive operations, then the increase in RT for axis-match (vs. axis-mismatch) sentences ought to be independent of whether the sentence contains a negative. The axis-match effect ought to be the same for false-affirmative and true-negative sentences.

In Exp. I, which served in some analyses as the control condition, the circle and sentence were presented simultaneously. To make the comparison easier, and to determine the effect of certain advance information, in Exp. II the sentence sign (IS, ISN'T) was presented for 3 sec. before the direction name and circle were shown, and in Exp. III the circle was presented for 3 sec. before the sentence was shown. To make the comparison more difficult, a memory load was imposed in Exp. IV.

Method

Apparatus.—Stimuli were presented on a Gerbrands Model T-3B-1 tachistoscope, which has six timers and three exposure fields (all set to maximum intensity: 4 mL.) each with a semiautomatic card changer. The RT was recorded by a Standard Electric Model S-1 timer (calibrated to .01 sec.) for Exp. I and II, and for the first 10 *S*s in Exp. III, and by a specially built electronic timer, in milliseconds, for the last 10 *S*s in Exp. III and for Exp. IV.

Subjects.—Twenty City College undergraduates served as paid *S*s in each experiment. Thus, a total of 80 *S*s served in the four experiments.

Stimulus materials.—All sentences and circles (capital O) were typed in uppercase in 12-point IBM Courier 72. In Exp. I and IV, each 4 × 6 in. stimulus card contained a two-word sentence (e.g., ISN'T NORTH, IS NORTH, IS EAST) and a circle, which was located either above, below, left of, or right of the sentence. In Exp. II, the sentence sign (IS, ISN'T) was typed on a separate card, which was shown for 3 sec. immediately preceding the stimulus card containing the direction name and circle. In Exp. III, the circle was typed on a separate card, which was shown for 3 sec. immediately preceding the stimulus card containing the sentence. The sentence was centered on the field, and the circle appeared 1 in. (Exp. I, IV), .75 in. (Exp. II), or .38 in. (Exp. III) from the sentence, either above, below, left, or right. The stimulus card was 31 in. from *S*.

Experimental design.—In each experiment, a within-*S*s design was used, except for one between-*S*s variable, hand assignment for true and false responses (half of the *S*s responded "true" with a left-hand key, half with a right-hand key). Each *S* received two replications of a basic set of 48 conditions, or 96 trials in all. The 48 represented all combinations of: (*a*) four direction names in the sentence, (*b*) two types of sentence form (affirmative, negative), (*c*) two degrees of truth value (true, false), and (*d*) three trials (which simply involved repetition of the same stimulus when the circle's location matched the direction name in the sentence; when circle and sentence did not match, each of the three nonmatching circle locations was used once each).

Procedure.—The *S*s were instructed to decide whether the sentence accurately described the location of the circle, and to respond as quickly as possible, but not at the expense of accuracy. For 10 *S*s in each experiment, the left key was pressed to signify a sentence–circle equivalence (true) and the right key a sentence–circle conflict (false), and for the other 10 *S*s the hand assignment was reversed.

The *S*s were told that a circle located above the sentence was accurately described by the sentences IS NORTH, ISN'T SOUTH, ISN'T EAST, and ISN'T WEST (true), but not by ISN'T NORTH, IS SOUTH, IS EAST, and IS WEST (false), and *S*s received eight practice trials.

A buzzer sounded when *S* made an error. All errors were replaced; the stimulus was repeated later in the experimental session until *S* responded correctly. All mean RTs reported below are based only on correct responses.

To initiate each trial, *S* pressed a foot pedal connected to a remote-start switch on the tachistoscope. The foot pedal was inoperative between trials. In Exp. I, after *E* said "go," *S* pressed the foot pedal, whereupon a blank field appeared for .3 sec., a fixation field for .7 sec., and the stimulus card for 2.0 sec. The *S*s were told to keep their eyes pointed at the fixation point, centered at the location of the sentence. In Exp. IV, *E* read five random digits to *S*, at a 2/sec rate, immediately before saying "go," and *S*, after responding to the sentence–circle combination, immediately wrote down as many of the five digits as he could in the proper order. The five-digit strings were devised by randomly permuting the digits 0 to 9, using the first five digits thus formed on one trial and the second five on a subsequent trial.

In Exp. II, when *S* pressed the foot pedal, the word IS or ISN'T appeared for 3.0 sec., followed for 2.5 sec. by the direction name and circle. In Exp. III, the circle appeared for 3.0 sec., followed for 2.5 sec. by the sentence. The RT was computed from the time of onset of the stimulus card that followed the 3-sec. preview of sentence sign (Exp. II) or circle (Exp. III).

Results and discussion

True negatives compared with all other sentence types.—The *F* ratios for RT for the four experiments combined and for each experiment analyzed separately are shown in Table 1. Mean overall RT was significantly shorter for affirmative sentences, true sentences, and vertical-axis sentences in which the direction name was NORTH or SOUTH. The horizontal–vertical difference was slightly larger overall for negative (77 msec.) than affirmative sentences (53 msec.), but only in Exp. IV did the Affirmative–Negative × Horizontal–Vertical interaction attain even marginal significance ($p < .10$).

Consistent with the additive model, the Affirmative–Negative × True–False inter-

TABLE 1

ANALYSIS OF VARIANCE OF MEAN RESPONSE TIME AND PERCENT ERRORS, BY SENTENCE TYPE

Source	*df*	*F* ratio				
		Exp. I–IV combined	Exp. I	Exp. II	Exp. III	Exp. IV
Response time						
Affirmative–Negative	1	276.68***	62.56***	120.19***	58.59***	69.77***
True–False	1	54.88***	8.74**	30.92***	19.43***	15.44***
Horizontal–Vertical	1	13.04***	7.72*	<1	1.72	6.27*
AN × TF	1	67.07***	32.63***	13.02**	178.69***	<1
AN × HV	1	2.00	<1	<1	<1	3.62
TF × HV	1	<1	<1	<1	<1	<1
AN × TF × HV	1	<1	<1	<1	<1	1.60
Errors						
Affirmative–Negative	1	43.62***	10.50**	1.71	19.44**	17.74***
True–False	1	<1	<1	1.85	2.65	1.50
Horizontal–Vertical	1	2.15	1.84	<1	<1	1.25
AN × TF	1	4.40*	<1	2.43	3.38	1.85
AN × HV	1	<1	1.65	<1	<1	<1
TF × HV	1	1.73	2.78	<1	2.42	3.60
AN × TF × HV	1	2.75	<1	<1	1.34	1.76

Note.—The error term for each source was its interaction with *S*s nested within hand-assignment groups. For the Exp. I–IV combined analysis, *S*s also were nested within groups across experiments. The error term had 72 degrees of freedom for the Exp. I–IV combined analysis, and 18 degrees of freedom for the analyses of the individual experiments.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

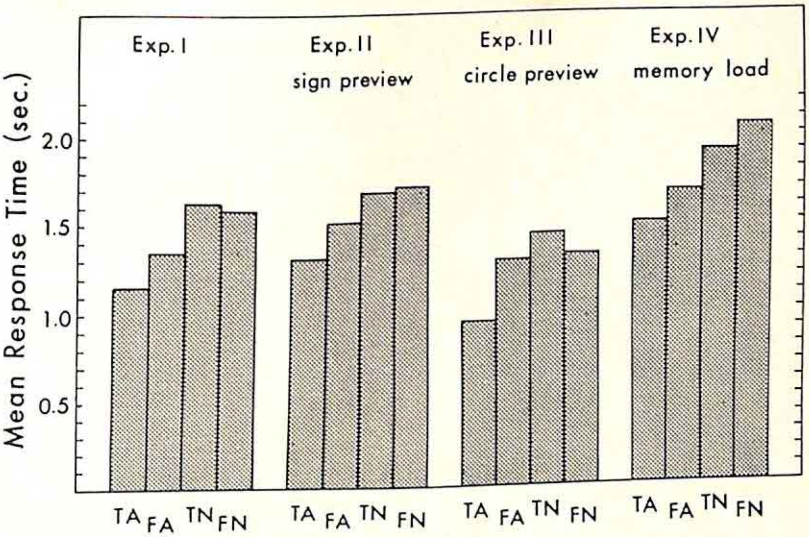


FIG. 1. Mean RT for true-affirmative (TA), false-affirmative (FA), true-negative (TN), and false-negative (FN) sentences, Exp. I–IV.

action was significant overall and in all experiments except Exp. IV. The Circle Preview (Exp. I vs. III) × Affirmative–Negative × True–False interaction was significant, $F\ (1,\ 36) = 19.67,\ p < .001$, as also was the Memory Load (Exp. I vs. IV) × Affirmative–Negative × True–False interaction, $F\ (1, 36) = 4.31, p < .05$). Contrary to the additive model, however, as Fig. 1 shows, the increment in RT (over the base-line true affirmatives) for true negatives, which involve two resettings of the response index, was less than the sum of the separate increments for false affirmatives and false negatives, which each involve one resetting of the response index. The shortfall from predicted RT for true negatives was 148 msec. in Exp. I, 224 msec. in Exp. II, 225 msec. in Exp. III, and 331 msec. in Exp. IV. In each experiment, the shortfall was significant ($p < .01$ or less), as indicated by Affirmative–Negative × Feature Match interactions obtained in special analyses in which the true–false variable was replaced by feature match versus mismatch (i.e., true affirmatives and false negatives vs. false affirmatives and true negatives).

The fact that the shortfall was reliable in Exp. III as well as the other experiments is important. In Exp. I, II, and IV, the sentence appeared simultaneously with, or partly preceded, the circle. Even though direction in the present case was nonbinary, *S* conceivably could have transformed a negative sentence into the set of equivalent affirmative sentences before attending to the circle. ISN'T SOUTH, for example, could be transformed into the set IS NORTH, IS EAST, and IS WEST, with each of these three affirmative sentences then being compared with the location of the circle. The removal of the negative would result in RT being shorter for true negatives than false negatives. Such negative-removing transformations do not occur, according to Trabasso et al. (1971), if the picture is presented before the sentence, as in Exp. III.

Mean RT was significantly higher in Exp. IV (memory load) than in Exp. I, $F\ (1,\ 36) = 14.75,\ p < .001$, significantly lower in Exp. III (circle preview) than Exp. I, $F\ (1,\ 36) = 10.53,\ p < .005$, and slightly higher in Exp. II (sentence sign preview) than in Exp. I, though only to a marginally significant extent, $F\ (1,\ 36) = 3.07,\ p < .10$. Neither sign preview (Exp. I vs. II), circle preview (Exp. I vs. III), nor memory load (Exp. I vs. IV) interacted significantly with affirmative–negative, true–false, or horizontal–vertical except that the Memory Load × True–False interaction was marginally significant, $F\ (1, 36) = 3.75, p < .10$. The direc-

TABLE 2

Mean Percent Errors, by Sentence Type

Sentence type	Exp. I–IV combined	Exp. I	Exp. II	Exp. III	Exp. IV
True affirmative	5.8	5.9	5.5	2.8	8.8
False affirmative	9.2	4.6	10.8	3.6	17.6
(Axis match)	(12.0)	(7.3)	(10.2)	(2.4)	(28.0)
(Axis mismatch)	(7.4)	(3.0)	(11.1)	(4.2)	(11.1)
True negative	17.8	15.8	11.4	15.8	28.2
(Axis match)	(20.0)	(16.9)	(12.1)	(15.8)	(35.0)
(Axis mismatch)	(16.6)	(15.3)	(11.1)	(15.8)	(24.2)
False negative	16.1	15.5	10.2	10.8	27.7

tion of the differences, however, generally conformed to the rule that the more difficult the task, the greater the affirmative–negative, true–false, or horizontal–vertical difference. For example, the horizontal–vertical difference was 153 msec. in Exp. IV, 58 msec. in Exp. I, 23 msec. in Exp. III, and 20 msec. in Exp. II.

It is surprising that *S*s did not benefit from the 3-sec. sign preview in Exp. II. The mean RT even was slightly longer in Exp. II than Exp. I. The longer RT in Exp. II may reflect the greater temporal uncertainty of the onset of the stimulus containing the direction name and circle (*S* had to wait 3.0 sec. rather than .7 sec.) and the fact that IS or ISN'T was in memory rather than in the visual display when *S* made his decision. More important, the affirmative–negative difference was not significantly lower in Exp. II (281 msec.) than in Exp. I (353 msec.). Negation coding normally may be performed after feature matching (Chase & Clark, in press). If negation coding is truly a separate operation, however, then, given the favorable circumstances in Exp. II, negation coding ought to have been accomplished before feature encoding and feature matching. Negation coding, with the resetting of the response index when ISN'T appeared, ought to have taken place during the 3-sec. preview. The difference in RT between negative and affirmative sentences ought to have been eliminated, but it was not. When questioned after the session, only 4 of the 20 *S*s reported that they sometimes reversed the hand assignment for true and false when an ISN'T was presented, thus making the requisite change in the response index. The *S*s may have needed practice to learn to use the sign preview; in Exp. II, however, the affirmative–negative difference was only slightly smaller in the second half of the session (257 msec.) than in the first half (305 msec.), $F < 1.0$.

Mean percent errors are shown in Table 2. The error rate was significantly higher in Exp. IV than Exp. I, t (38) = 3.61, $p < .001$, but the error rates in Exp. II and Exp. III did not differ significantly from that in Exp. I. The F ratios on errors in Table 1 show a significant affirmative–negative main effect and, for all experiments combined, a significant Affirmative–Negative × True–False interaction. In all experiments the error rate was the highest for the true negatives (see Table 2). The higher error rate for true negatives may reflect a greater guessing rate for that sentence type, but, more likely, it may merely be due to the fact that a true negative presents a greater number of choice points, at any one of which a misstep could produce an error in *S*'s response. If many errors were fast guesses, then many of the "correct" RTs also would be fast guesses. Even so, however, the increase in error rate for true negatives was not large enough to explain the considerable shortfall from predicted RT for true negatives. Further, there is no evidence that error responses were made more rapidly than correct responses. For the 30 *S*s in Exp. III and IV for whom error RTs were recorded, RT was slightly and insignificantly longer (18 msec., $t < 1$) for errors than for the correct responses that replaced the errors.

In Exp. IV, the difficulty of the circle–sentence comparison task had little effect on *S*'s memory for the five-digit string which *S* received immediately before the circle–sentence stimulus and had to recall immediately after making his judgment of the circle–sentence combination. The mean number of digits not correctly recalled in the proper position in the five-digit string was .8 for true affirmatives, .9 for false affirmatives, .9 for true negatives, and 1.0 for false negatives. The true–false difference was significant, F (1, 18) = 5.16, $p < .05$, and the affirmative–negative dif-

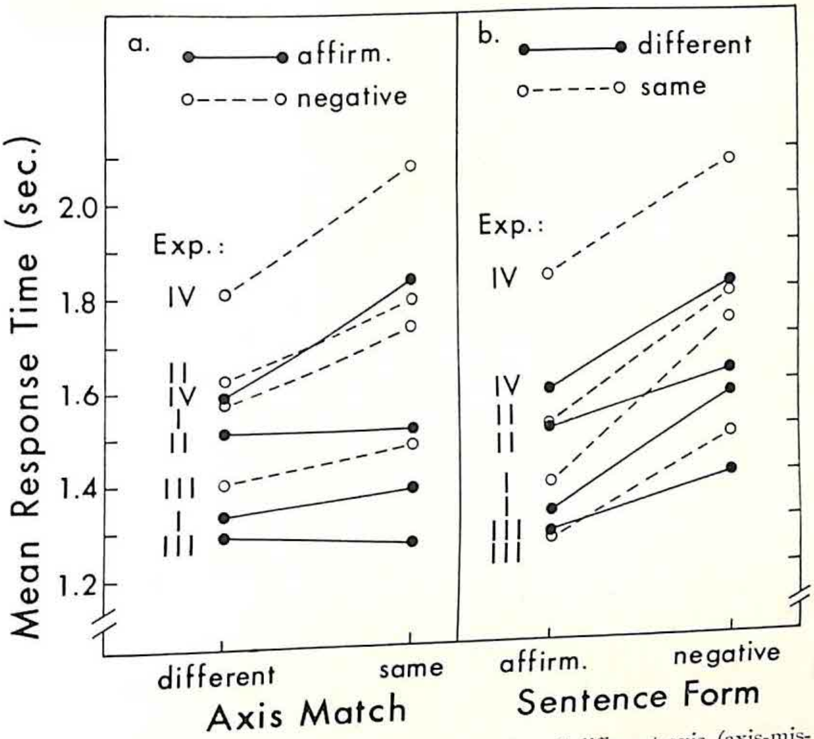


FIG. 2. Mean RT for same-axis (axis-match) and different-axis (axis-mismatch) feature mismatches, for affirmative and negative sentences, Exp. I–IV. (Only data for false-affirmative and true-negative sentences are shown. The same data are plotted in the left panel, a, as in the right panel, b.)

ference was marginally significant, $F (1, 18) = 3.93$, $p < .10$.

True negatives compared only with false affirmatives.—A comparison of the RT of false affirmatives with true negatives provides a second test of the additive model. The affirmative–negative and true–false dimensions are confounded in the comparison of false affirmatives with true negatives, but for the present analysis it will be assumed that only the affirmative–negative difference was effective. Two facts justify the disregard of the true–false component. First, as Fig. 2b shows, RT was consistenly and considerably higher for true negatives than false affirmatives. Second, because direction is nonbinary, taking one of four rather than two values, a negative-removing transformation cannot be applied before the feature-match operation. According to the additive model, then, true negatives and false affirmatives would both involve a resetting of the response index due to a feature mismatch, and would differ only in that true negatives also would involve a resetting due to a negative.

A feature mismatch between sentence and circle can involve either an axis match, where the features retain the same axis (horizontal or vertical) but differ on specific directions, or an axis mismatch, where both axis and specific direction differ. As Fig. 2a shows, RT is longer in the axis-match (same-axis) case than in the axis-mismatch (different-axis) case. Apparently, feature matching is more difficult in the axis-match case. The F ratios for RT in Table 3 show that both the axis-match and affirmative–negative main effects were significant ($p < .001$).

According to the additive model, feature matching and negation coding involve separate operations and should not interact. Contrary to the additive model, however, the Axis Match × Affirmative–Negative interaction for the four experiments combined was significant ($p < .01$). The axis-match effect was 92 msec. greater for negative

TABLE 3

Analysis of Variance of Mean Response Time and Percent Errors, by Sentence Type (False Affirmative and True Negative), for Feature-Mismatch Sentences

Source	*df*	*F* ratio				
		Exp. I–IV combined	Exp. I	Exp. II	Exp. III	Exp. IV
Response time						
Axis match	1	31.78***	17.92***	8.37**	1.54	13.56**
Affirmative–Negative[a]	1	88.59***	48.22***	54.25***	15.53***	14.41**
Horizontal–Vertical	1	8.86**	5.22*	<1	<1	4.92*
AM × AN	1	10.23**	3.51	19.40***	3.04	<1
AM × HV	1	2.68	5.58*	<1	1.41	2.05
AN × HV	1	2.64	1.01	<1	<1	2.40
AM × AN × HV	1	<1	<1	1.72	<1	<1
Errors						
Axis match	1	6.25*	<1	<1	<1	7.18*
Affirmative–Negative[a]	1	20.04***	9.34**	<1	12.72**	5.09*
Horizontal–Vertical	1	<1	1.41	<1	<1	<1
AM × AN	1	<1	<1	<1	<1	<1
AM × HV	1	1.83	<1	2.25	2.16	2.60
AN × HV	1	<1	3.10	<1	1.96	<1
AM × AN × HV	1	<1	<1	2.03	<1	<1

Note.—The error term for each source was its interaction with *S*s nested within hand-assignment groups. (For the Exp. I–IV combined analysis, *S*s also were nested within groups across experiments. The error term had 72 degrees of freedom for the Exp. I–IV combined analysis, and 18 degrees of freedom for the analyses of the individual experiments.)

[a] The affirmative–negative dimension actually was confounded with the true–false dimension.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

(166 msec.) than affirmative sentences (74 msec.). Every experiment showed the same pattern of results. The 157-msec. increase in the axis-match effect for negative sentences in Exp. II was significant, $p < .001$, the 94-msec. increase in Exp. I and the 101-msec. increase in Exp. III were marginally significant, $p < .10$, and the 19-msec. increase in Exp. IV was not significant. It is important that the result was of about the same magnitude in Exp. III as in the other experiments. The circle was presented first, before the sentence, in Exp. III, making it even less likely that *S*s removed the negative through a sentence transformation before comparing the sentence and circle representations. On error rate, the Axis Match × Affirmative–Negative interaction was not significant, as the *F* ratios in Table 3 indicate.

Although the Axis Match × Horizontal–Vertical and Affirmative–Negative × Horizontal–Vertical interactions were not significant, in both cases, as in the case of Axis Match × Affirmative–Negative, the interaction was "positive." The horizontal–vertical difference was larger for axis-match (98 msec.) than axis-mismatch sentences (35 msec.) and larger for negative (82 msec.) than affirmative sentences (44 msec.).

General Discussion

The present data disconfirm the additive model in two respects: (*a*) RT for true negatives was lower than predicted, and (*b*) the axis-match effect was larger when the sentence contained a negative. The results in Fig. 1 suggest instead a parallel-processing model, in which feature matching and negation coding occur simultaneously and independently of each other. If processing is parallel, then the time needed to perform both operations ought to equal the time needed for the slower of the two operations. Mean overall RT was longer for false negatives than false affirmatives, indicating that it took longer to handle a negative than a feature mismatch, but was virtually the same for true negatives (1,659 msec.) and false negatives (1,664 msec.), indicating that it took no longer to handle a negative and a feature mismatch than to handle a negative alone.

The Axis Match × Affirmative–Negative interaction, however, indicates that, contrary to the parallel model, feature matching and negation coding are not truly independent operations. The parallel model may be modified to be consistent with the data, however, as shown by the parallel-dependent model proposed by Moray (1967) (cf. also the constant-energy parallel model of Townsend, 1971). In the parallel-dependent model, several simultaneous operations share a limited-capacity central processor, and independent parallel processing is possible when the processor is not overloaded.

It is proposed here that all operations performed during a circle–sentence comparison, including the holding of items in short-term store, etc., are parallel dependent, and that, past some point, adding new operations to the task will cause all operations to be performed more slowly (cf. Meyer, in press). In addition, the encoding and comparison operations may operate in a parallel-dependent rather than serial manner. After having encoded one item (sentence or picture), and while encoding the second item, *S* may concurrently be making test comparisons between the fully encoded first item and the partially encoded second item. Further, if the first item is more difficult to encode (e.g., a sentence with a negative), it also may be more difficult to keep in storage, and thus may slow the encoding of the second item in addition to making the comparison process more difficult. Meyer (in press) had his *S*s compare sentences with information in memory, and he reported results that "support the view that sentence negation requires extra processing capacity and thereby reduces the rate of searching long-term memory." In the present case, the axis-match effect may reflect increased difficulty at the feature-encoding rather than the comparison (i.e., feature-matching) stage, and negation may accentuate the axis-match effect because of the greater processing demands that negative sentences make at the encoding and storage rather than comparison stage. Regardless of where negation affects processing time, though, the present findings, as well as those of Meyer, do clearly indicate that negation is not always handled in a separate, additive operation.

The effect of both a negative and an axis match on RT is larger when overall RT is longer (see Fig. 2). Rather than bearing a special relationship with each other, feature matching and negation coding apparently affect each other simply because both are influenced by any factor that increases the demands on the central processor, whether it be negation, an axis match, a memory load, a sentence containing EAST or WEST (horizontal axis) rather than NORTH or SOUTH (vertical axis), or the simultaneous (vs. successive) presentation of sentence and circle. The effect of task difficulty on negation (Fig. 2b) is less noticeable than that on axis match (Fig. 2a). Perhaps negation involves lengthy serial operations that always lead to a longer RT, even when the central processor is not overloaded.

The effect of task difficulty on feature matching, on the other hand, is so striking as to suggest a switch in mode of processing: from exhaustive feature matching, when the task is easy, to self-terminating feature matching, when the task is difficult. Self-terminating feature matching means the following: *S* first compares the axis of the direction name in the sentence with the axis of the circle. If an axis mismatch is encountered, feature matching self-terminates immediately. If an axis match is encountered, *S* checks further to determine whether the specific directions within the axis are the same or different. The check further may require more encoding as well, if the circle or sentence or both had only been partially encoded, to the level of axis value or semantic marker when first compared. In exhaustive feature matching, on the other hand, the further comparison involving specific directions would be carried through even when not necessary, i.e., when an axis mismatch had been encountered. When the task is easy, the availability of unused processing capacity may enable the extra operations entailed by exhaustive feature matching to be readily and quickly performed, perhaps in parallel with each other. When the task is difficult, on the other hand, *S*s may make self-terminating matches whenever possible so as to free processing capacity for other operations, and thus avoid lengthening RT.

P. Liss (personal communication, November, 1971) has suggested an elaboration of the parallel-dependent model. In an unpublished experiment, Liss presented a varying number of items for *S* to store for subsequent comparison with a test item. When the task was easy, memory search was exhaustive, as Sternberg

(1966) had found. When more difficult geometrical figures were presented, however, memory search was self-terminating. Slowing down memory search seemingly gave *S* a greater opportunity to terminate memory search immediately upon encountering a match between a memory item and the test item. In the present study, according to Liss, feature matching may have changed from exhaustive to self-terminating as the task became more difficult because as feature matching slowed down under the more difficult conditions, it offered a better target for self-termination. When the task was easy, feature matching ran itself off too rapidly to allow self-termination.

REFERENCES

Chase, W. G., & Clark, H. H. Mental operations in the comparison of sentences and pictures. In L. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, in press.

Meyer, D. E. Verifying affirmative and negative propositions: Effects of negation on memory retrieval. In S. Kornblum (Ed.), *Attention and Performance IV*. New York: Academic Press, in press.

Moray, N. Where is capacity limited? A survey and a model. *Acta Psychologica*, 1967, **27**, 84–92.

Sternberg, S. High-speed scanning in human memory. *Science*, 1966, 153, 652–654.

Sternberg, S. The discovery of processing stages: Extensions of Donder's method. *Acta Psychologica*, Amsterdam, 1969, **30**, 276–315.

Townsend, J. T. A note on the identifiability of parallel and serial processes. *Perception & Psychophysics*, 1971, **10**, 161–163.

Trabasso, T., Rollins, H. & Shaughnessy, E. Storage and verification stages in processing concepts. *Cognitive Psychology*, 1971, **2**, 239–289.

(Received November 24, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 285–289

STIMULUS MODALITY EFFECTS ON FORGETTING IN SHORT-TERM MEMORY[1]

DON L. SCARBOROUGH[2]

Brooklyn College of the City University of New York

Retention of visually and of auditorily presented trigrams (CCC) was compared using the Peterson distractor technique. In contrast to previous experiments, retention of visually presented trigrams was superior, particularly at the longer retention intervals, e.g., 18 sec. These results extend recent evidence that the retention of visually presented displays of symbols and of auditorily presented sequences of symbols can be mediated by different memories that are largely independent of each other. That memory for visually presented stimuli is less disrupted by the interpolated task (counting backward) than is memory for auditorily presented stimuli follows from the hypothesized independence of the two memory systems.

Peterson and Peterson (1959) reported rapid forgetting of three-consonant trigrams if a task of counting backward by threes or fours was inserted between presentation and recall of a trigram. This is a reliable phenomenon and has been replicated qualitatively by other investigators (e.g., Keppel & Underwood, 1962; Melton, 1963; Murdock, 1961). However, there appear to be differences between the replications in the amount of forgetting produced by the interpolated task. For example, Peterson and Peterson reported that after 18 sec. of the interpolated task, the mean probability of correct trigram report was less than .1 (Exp. I). Similarly, Murdock (1961) reported a .16 probability of correct report after 18 sec. of the interpolated counting task (Exp. I). On the other hand, Keppel and Underwood reported approximately a .45 probability of correct report after 18 sec. of interpolated counting (Exp. 2 and 3), and Melton reported a .51 probability of correct report after 32 sec. (his longest retention interval) of interpolated counting. Thus, though all *E*s have found considerable forgetting of the trigrams following the interpolated task, there are large disparities in the amounts of forgetting observed. However, the procedural details of the four studies differed in several respects. The present study focuses on one dimension in which the studies differed: visual versus auditory trigram presentation. On the one hand, Peterson and Peterson and Murdock (1961) presented the stimuli auditorily while Keppel and Underwood (1962, Exp. 2 and 3) and Melton presented the stimuli visually. It may be that visual stimulus presentation leads to a memory code that is less affected by counting backward than is a memory code produced by auditory trigram presentation. The experiment reported here was designed to explicitly compare visual and auditory trigram presentation under conditions similar to the original study by Peterson and Peterson.

METHOD

There were three experimental groups of six undergraduate paid volunteers. Each *S*, run individually, received practice trials and two blocks of test trials in a 1-hr. session. The practice trials gave *S*s experience in counting backward. For the practice trials and the first block of test trials, the three groups differed with respect to trigram presentation modality. For one group, the trigrams were presented visually (V). For a second group, the trigrams were presented auditorily (A), while for the third group, the trigrams were simultaneously presented both visually and auditorily (A + V). In the second block of test trials, the trigrams were presented auditorily for all groups so that the three groups could be compared on a common task. This common task was presented last because it was thought that if auditory trigram presentation were used in the first test block, *S*s might later persist in using auditory coding when the presentation modality was changed.

---

[1] This research was supported by a Faculty Research Award from the City University of New York. Mary V. Justice helped in the preparation of the stimuli and in the analysis of the data. I wish to thank Doris Aaronson, Deborah Macmillan, and Hollis Shapiro for many helpful comments on earlier drafts of this paper.

[2] Requests for reprints should be sent to Don L. Scarborough, Department of Psychology, Brooklyn College of the City University of New York, Brooklyn, N. Y. 11210.

Stimuli.—Two lists of 36 consonant trigrams were constructed from Witmer's list of trigrams with 0% to 17% association value (Hilgard, 1951). The lists were constructed so that, in keeping with the procedure employed by Peterson and Peterson (1959), no two successive trigrams contained letters in common. A three-digit number drawn from a table of random numbers was paired with each trigram. The two lists of trigram-number pairs were tape recorded with a warning tone (1,000 Hz., .5 sec.). The onset of the tone preceded the trigram by 1 sec. Similarly, the trigram onset preceded the three-digit number by 1 sec. The duration of the trigrams and of the numbers was approximately .7 sec. as measured with an oscilloscope. This tape was used for A and A + V test trials. For V test trials, the trigrams were deleted from a copy of this tape, and a 1,000-Hz. tone, recorded in place of each trigram on the second track of the tape, served to initiate presentation of a visual trigram on a projection tachistoscope.

Two practice tapes were prepared in the same manner. These consisted of 18 trials with the same trigram but a different three-digit number on each trial.

For V and A + V test trials, slides were made of the trigrams, using photographs of standard dry-transfer letters in a Futura Demi-Bold font. The letters appeared dark on a light background when projected. The visual angle subtended by the display was approximately 2° horizontally by 35′ vertically.

The duration of the interpolated counting backward task varied from trial to trial. The six durations used were 0, 3, 6, 9, 12, and 18 sec., measured from the end of the trigram presentation to the recall signal. The order of the six intervals was determined by a 6 × 6 Latin square. Three such Latin squares were constructed and each Latin square was presented to two *S*s within each experimental group for the first block of test trials. In the first block of test trials, Stimulus List 1 was presented to one of each pair of *S*s within an experimental group receiving the same order of interpolated task durations, and Stimulus List 2 was presented to the other *S*. This was reversed for the second block of test trials.

Procedure.—The *S*s were randomly assigned to one of the three experimental groups upon arrival in the laboratory. The *S* was seated in a soundproof experimental chamber where he was read the instructions appropriate to his group. Following this, he practiced for 18 trials on counting backward by three's. The same trigram was presented on every practice trial, but the three-digit number was changed. The trigram was presented in the same modality (or modalities) to be employed in the first block of test trials. Each of the six delay intervals was presented three times in the practice trials.

On each trial in a block, two yellow lights on a panel in front of *S* turned on to warn *S* that the trial was about to begin. Approximately 5 sec. later, *S* heard the 1,000-Hz. warning tone and the yellow lights turned off. The trigram was presented 1 sec. later over high-quality earphones and/or for 1 sec. on a rear-projection screen illuminated by a two-field projection tachistoscope with blank pre- and postdisplay fields. One second after trigram presentation, *S* heard a three-digit number on the earphones, except on the 0-sec. delay trials. The *S* was instructed to repeat the number aloud and begin counting backward by three's from that number in time with audible 1/sec metronome clicks. At the end of the delay interval, the metronome stopped and two green recall lights turned on for a 10-sec. recall interval. Following the recall period, the green lights turned off and the yellow warning lights for the next trial turned on. Every six trials, there was a pause until *S* was ready to proceed. The *S*s were given a rest after the practice trials and between the two blocks of test trials. All responses were tape-recorded.

RESULTS

The mean probabilities of correctly reporting the trigrams in the first and second blocks of test trials are shown in Fig. 1. Trigram report was also analyzed in terms of the mean number of letters correctly reported on each trial. As the results were very similar to the effects observed with correct trigram report, only the latter data are reported. The data from each block were analyzed separately. For the first test block, the arc-sine transformation (Winer, 1962) of the probabilities for each *S* was analyzed in a three-factor analysis of variance (groups, interpolated task duration, and *S*s), with *S*s as a nested factor within groups. The effect of groups, $F(2, 15) = 15.3$, $p < .01$, and of interpolated task duration, $F(5, 75) = 35.0$, $p < .01$, were both significant. The Groups × Interpolated Task Duration interaction did not quite reach significance, $F(10, 75) = 1.8$, $p < .10$. The error term for the test of this interaction was the interaction of *S*s with duration. The interaction with *S*s was larger for the A + V group than for either the A group or the V group, though the difference was not significant when tested by Cochran's test for homogeneity of variance (Winer, 1962). However, the larger variation of the error term for the A + V group ac-

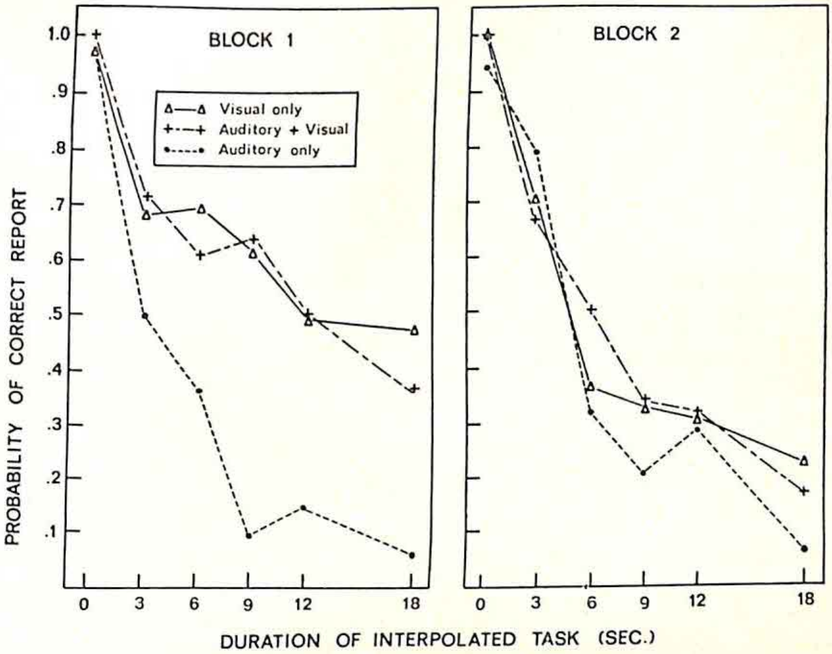


FIG. 1. Probability of correct trigram report for each of two blocks of test trials. (The experimental groups differed with respect to trigram presentation modality only in Block 1. All groups received auditory trigram presentation in Block 2.)

counts for the failure of the Groups × Interpolated Task Duration interaction to reach significance. Comparing only the A and the V groups, the interaction was significant, $F\ (5, 50) = 3.6$, $p < .01$, as were the main effects. The same comparison between the A and the A + V groups again yielded significant main effects, but the interaction was not significant, $F\ (5, 50) = 1.9$, $p > .10$.

Trigrams were presented auditorily for all three groups in the second test block. An analysis of variance on the arc-sine transformation of the probabilities of correct trigram report showed a significant effect of interpolated task duration, $F\ (5, 75) = 52.6$, $p < .01$. However, neither the effect of experimental group, $F\ (2, 15) = .9$, $p > .25$, nor the Groups × Interpolated Task Duration interaction, $F\ (10, 75) = .9$, $p > .25$, approached significance.

To test whether the groups differed in performance of the interpolated task, the mean number of times each *S* counted backward on a trial averaged across all interpolated task durations was analyzed by an analysis of variance with the factors of groups, *S*s, and blocks (Block 1 vs. 2). The results showed no significant effect of groups, $F\ (2, 15) = .1$, $p > .25$, blocks, $F\ (1, 15) = .1$, $p > .25$, or the Groups × Blocks interaction, $F\ (2, 15) = 1.1$, $p > .25$.

DISCUSSION

It is quite clear that visual trigram presentation (as compared with auditory trigram presentation) led to superior retention in this experiment. The reliability of this effect is supported by the similarity of the present data to the results of previous experiments. Recall performance for the A group on the first block of trials is reasonably close to (though somewhat lower than) the levels of performance reported by Peterson and Peterson (1959), Murdock (1961), and Keppel and Underwood (1962, Exp. 1), for equivalent delays. Performance by the two groups with visual trigram presentation (V and A + V) is similar to the performance levels reported by Keppel and Underwood (1962, Exp. 2) and Melton (1963) where visual trigram presentation was also used. Thus, in this task, retention was better for visually presented stimuli than for auditorily presented stimuli. One possible explanation of this difference is that visual presentation may have allowed more time for trigram rehearsal (e.g., because of faster perception). However, in one experiment, Peterson and Peterson permitted 3 sec. of silent or vocal trigram rehearsal before the interpolated task. The probabilities of correct trigram report for 9 and 18 sec. of interpolated counting were still lower than the probabilities for the same delays in the present experiment with visual trigram presentation. This indicates that a 1-sec. visual trigram presentation produced better retention than auditory trigram presentation with 3 sec. of rehearsal.

Several *E*s (Brooks, 1967, 1968; Kroll, Parks, Parkinson, Bieber, & Johnson, 1970; Murdock, 1971; Scarborough, 1972) have proposed that short-term retention of visually presented information can be mediated by a memory system other than the auditory short-term storage system studied by Conrad (1962, 1964), Sperling (Sperling, 1960; Sperling & Speelman, 1970), and others. The results of the present experiment further support this hypothesis. With auditory trigram presentation, it seems likely that *S*s tried to retain the trigrams in an auditory short-term store that was disrupted by performance of the interpolated task. On the other hand, visually presented trigrams evidently were stored in a form that was much less disrupted by the verbal interpolated task. Alternatively, it seems likely that the superiority of visual trigram presentation would have disappeared if the interpolated task had required the processing of visual information. This hypothesis is supported by the results of Morton and Holloway (1970), who reported that adding a stimulus suffix and response prefix had a large deleterious effect on report of auditory or visual sequences of digits or letters only when the suffix and prefix were in the same modality as the stimulus sequence. Also, Tell (1971) found that binary manual responses in an interpolated task interfered less with retention of verbalized trigrams than did equivalent binary verbal responses.

The hypothesis that auditory and visual stimulus presentations lead to different forms of storage appears to provide an explanation for the release from proactive interference (PI) that is sometimes observed when the stimulus presentation modality is changed within a series of short-term memory trials (Wickens, 1970). However, Hopkins, Edwards, and Gavelek (1971) in an investigation of this effect reported that although they found a release from PI when there was a change from visual to auditory stimulus presentation, they did not find such an effect when there was a change from auditory to visual stimulus presentation. Such a result appears to be inconsistent with the hypothesis proposed in this paper. A closer inspection of their data suggests a different interpretation. In general, their results indicate that retention of auditory stimuli was superior to retention of visual stimuli, possibly because their interpolated task required *S*s to add a sequence of 14 pairs of *visually* presented digits. As a result, a switch to auditory stimulus presentation after several trials of visual stimulus presentation should produce an improvement in performance for two reasons: (*a*) release from PI because the auditory stimuli are stored differently and (*b*) superior overall performance with auditory stimuli. On the other hand, a switch to visual stimulus presentation after several trials of auditory stimulus presentation should produce equivocal results because: (*a*) performance should improve because of the release from PI and (*b*) performance should get worse because of generally poorer performance with visually presented stimuli in their experiment. In fact, their results show a large improvement in performance with a shift from visual to auditory presentation, and practically no improvement with a shift from auditory to visual presentation.

Other *E*s have compared visual and auditory trigram presentation and have reported superior performance with auditory trigram presentation. Cooley and McNulty (1967) report that using the Peterson and Peterson (1959) procedure, retention was best for auditorily presented trigrams at short retention intervals (3 and 6 sec.) although for longer retention intervals (e.g., 18 sec.) there was little difference between auditory and visual presentation. Similarly, Grant and McCormack (1969) compared retention of auditorily and visually presented trigrams when the letters were presented sequentially for both modalities. As with Cooley and McNulty, they conclude that auditory presentation is superior to visual presentation, especially at shorter retention intervals, e.g., 0 to 6 sec. This conclusion seems unjustified in the case of both experiments, for the data show that report of the auditorily presented trigrams was better than report of visually presented trigrams with a 0-

sec. retention interval. The inability of *S*s to report the visual trigrams correctly immediately after presentation indicates that the visual trigrams may not have been accurately perceived at the time that the interpolated task began. It is clear that under these conditions, performance at the longer retention intervals cannot be meaningfully compared with respect to whether visual or auditory presentation produces better retention.

In summary, in contrast to the widely held belief that both visually and auditorily presented stimuli are retained in a common auditory short-term store, this experiment supports the hypothesis that visually presented linguistic stimuli (trigrams) can be retained in a different memory system. Comparison of the present study with other experiments indicates that further research comparing modalities should give more attention to the level of original learning for each modality, and to the information-processing demands of the interpolated tasks.

REFERENCES

BROOKS, L. R. The suppression of visualization by reading. *Quarterly Journal of Experimental Psychology*, 1967, **19**, 289–299.

BROOKS, L. R. Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, 1968, **22**, 349–368.

CONRAD, R. An association between memory errors and errors due to acoustic masking of speech. *Nature*, 1962, **193**, 1314–1315.

CONRAD, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75–84.

COOLEY, R. K., & MCNULTY, J. A. Recall of individual CCC trigrams over short intervals of time as a function of mode of presentation. *Psychonomic Science*, 1967, **9**, 543–544.

GRANT, K. W., & MCCORMACK, P. D. Auditory and visual short-term memory with successive presentation in both modalities. *Psychonomic Science*, 1969, **17**, 341–342.

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

HOPKINS, R. H., EDWARDS, R. E., & GAVELEK, J. R. Presentation modality as an encoding variable in short-term memory. *Journal of Experimental Psychology*, 1971, **90**, 319–325.

KEPPEL, G., & UNDERWOOD, B. J. Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 153–161.

KROLL, N. E. A., PARKS, T., PARKINSON, S. R., BIEBER, S. L., & JOHNSON, A. L. Short-term memory while shadowing: Recall of visually and of aurally presented letters. *Journal of Experimental Psychology*, 1970, **85**, 220–224.

MELTON, A. W. Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 1–21.

MORTON, J., & HOLLOWAY, C. M. Absence of a cross-modal "suffix effect" in short-term memory. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 167–176.

MURDOCK, B. B., JR. The retention of individual items. *Journal of Experimental Psychology*, 1961, **62**, 618–625.

MURDOCK, B. B., JR. Four channel effects in short-term memory. *Psychonomic Science*, 1971, **22**, 197.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

SCARBOROUGH, D. L. Memory for brief visual displays of symbols. *Cognitive Psychology*, 1972, **3**, (in press).

SPERLING, G. The information available in brief visual presentations. *Psychological Monographs*, 1960, **74** (11, Whole No. 498).

SPERLING, G., & SPEELMAN, R. G. Acoustic similarity and auditory short-term memory: experiments and a model. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

TELL, P. M. Influence of vocalization on short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 149–156.

WICKENS, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received December 10, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 290–294

INTERAURAL ATTENTION SHIFTING AS RESPONSE[1]

LAWRENCE T. GUZY[2] AND SEYMOUR AXELROD

State University of New York at Buffalo

Fifteen college students were instructed to count the number of regularly repeated clicks, presented either to one ear or in alternation to the two ears, at repetition rates from 2/sec to 8/sec in trains varying in length from 2 to 20 clicks. Relative to monotic clicks, dichotically alternating clicks were undercounted, the degree of relative undercounting being a direct function of number of clicks/train and repetition rate. The results are consonant with a view of repetitive shifting of attention between the ears as at least formally similar to a repetitive motor response, and perhaps as dependent upon oscillating side-to-side orienting movements, e.g., of the eyes.

Repetitive clicks alternately delivered to the two ears are perceived as occurring more slowly than clicks delivered to a single ear, at click-repetition rates between about 3/sec and 40/sec, and with train durations greater than about 3 sec. The apparent slowing has been interpreted as indicating difficulty in integrating dichotically alternating inputs into a single percept and attributed to inability to shift attention between the ears rapidly and repetitively (Axelrod & Guzy, 1968; Axelrod, Guzy, & Diamond, 1968; Guzy & Axelrod, 1971).

The study reported here had its genesis in *E*s' observations that (*a*) rapid dichotic click trains appeared to slow down only after a few clicks had been presented, and (*b*) the slowing appeared earlier in a faster train (e.g., 10/sec) than in a slower one (e.g., 5/sec). This is, of course, just what would be expected if interaural attention shifting were analogous to a paced, repetitive, fatigable motor response in that it can be performed rapidly a few times, slowly many times, but not rapidly many times. (Reasons for thinking that there may be more than an analogy here are presented in the Discussion below.) A prediction which follows from such a view is that the number of rapidly alternating clicks the listener can perceive is an inverse function of both the number of clicks in the train and the presentation rate; i.e., at a presentation rate in which a long train of alternating clicks would be perceived imperfectly, perception of a small number of clicks is possible. Within the range of repetition rates in which some degree of dichotic slowing occurs with long trains, (*a*) adequacy of processing should break down as the train is extended, and (*b*) the breakdown should occur earlier in the train as the repetition rate is increased.

Test of the predictions obviously requires a task in which the listener must begin processing the stimuli at the onset of the train. We therefore required *S*s to count the clicks in monotic and dichotic trains of eight lengths between 2 and 20 clicks, with intratrain click-repetition rates of 2, 4, 6, and 8/sec.

METHOD

Subjects.—The *S*s were 15 men and women whose participation fulfilled a course requirement in introductory psychology. They ranged in age from 17 to 26 yr., the mean age being 19.8. None of them admitted awareness of any hearing deficits.

Apparatus.—Clicks were generated by .1-msec., 20-v. pulses from Tektronix (Type 161) pulse generators, triggered by Tektronix (Type 162) waveform generators, and fed into a Clevite "Brush" high-impedance crystal headset. Pulses were monitored as to frequency with a Hickok (DP-150) digital meter, and as to duration and amplitude with a

---

[1] This work was supported by a research grant from the National Institute of Neurological Diseases and Stroke (NS-06862), an institutional grant from the National Science Foundation, and Research Scientist Development Award K3-MH-5787 from the National Institute of Mental Health, all to the second author. The results were reported in a paper presented at the meeting of the Eastern Psychological Association, New York City, April 1971.

[2] Requests for reprints should be sent to L. T. Guzy, who is now at Department of Education and Psychology, Alliance College, Cambridge Springs, Pennsylvania 16403.

TABLE 1

Mean Number of Clicks Counted

Repetition rate in clicks/sec	Mode	No. of clicks presented							
		2	3	4	5	6	8	10	20
2	Monotic	2.00 (.000)	3.02 (.041)	4.01 (.035)	5.01 (.035)	6.02 (.041)	8.03 (.062)	9.99 (.026)	20.00 (.000)
	Dichotic	2.00 (.000)	3.01 (.026)	4.00 (.000)	5.01 (.084)	6.01 (.027)	7.99 (.026)	10.00 (.038)	20.00 (.080)
4	Monotic	2.02 (.056)	3.06 (.099)	4.04 (.091)	4.93 (.296)	6.01 (.164)	7.97 (.110)	10.02 (.237)	19.80 (.327)
	Dichotic	2.00 (.000)	3.02 (.056)	3.85 (.307)	4.84 (.422)	5.78 (.570)	7.55 (.802)	9.33 (1.063)	17.45 (2.630)
6	Monotic	2.02 (.056)	3.05 (.092)	3.98 (.132)	4.85 (.200)	5.63 (.375)	7.48 (.677)	9.13 (.763)	17.16 (1.358)
	Dichotic	1.99 (.070)	2.75 (.304)	3.51 (.497)	4.19 (.574)	4.69 (.565)	5.99 (.851)	7.18 (1.101)	13.23 (1.773)
8	Monotic	2.02 (.056)	3.02 (.054)	3.81 (.210)	4.66 (.282)	5.40 (.374)	6.79 (.592)	8.21 (.789)	15.32 (1.430)
	Dichotic	2.01 (.122)	2.59 (.364)	3.16 (.460)	3.78 (.488)	4.39 (.527)	5.34 (.660)	6.45 (.779)	11.87 (1.299)

Note.—Standard deviations are shown in parentheses.

Tektronix (Type 360) indicator-unit oscilloscope. Number of clicks presented was predetermined by appropriate setting of a Hunter (111-C) silent decade interval timer and verified by recording stimuli graphically on a Medcraft (Model D) electroencephalograph. At a stimulus repetition rate of 40/sec and with the earphones held tightly against the microphone of a General Radio (Type 1555A) Sound Survey Meter, a steady recording of 58 db. re .0002 microbar was obtained from each of the earphones.

Procedure.—The *S*s were run individually. On each trial, a train of clicks was delivered via one (monotic) or both earphones (dichotic alternation). The *S*, seated at a table, was instructed to count the clicks as they occurred and report to *E* the number of clicks counted *immediately* after the trial. The *S* was discouraged from estimating the number presented, regardless of the difficulty in counting. Special emphasis was made to insure that *S* understood that he was to take into account both ears when the clicks were presented to both ears in dichotic alternation.

In the experimental run, each *S* received 640 trials arranged in 80 blocks of 8 trials each. For each block, mode (dichotic or monotic) and repetition rate (2, 4, 6, or 8/sec) were fixed; and the number of clicks presented (2, 3, 4, 5, 6, 8, 10, and 20) varied from trial to trial randomly without replacement. For each block, mode was assigned quasi-randomly with the restrictions that (*a*) no more than 2 successive blocks were run under the same mode, and (*b*) 40 blocks each were run under monotic and dichotic modes. Of the 40 monotic blocks, 20 were left-ear blocks and 20 right-ear blocks. For each block, repetition rate was assigned quasi-randomly, with the restrictions that (*a*) no 2 successive blocks were run with the same rate, and (*b*) 20 blocks were run with each of the four rates. Thus, each *S* received 10 trials for each Mode × Rate × Number cell; the mean of these 10 trials was taken as his score for that cell. Before the experimental run, in order to insure that *S* understood the task, 20 practice trials were given (10 monotic and 10 dichotic) with a slow repetition rate (1/sec), and with the number of clicks/train ranging from 2 through 10. Responses were invariably veridical during these practice trials.

RESULTS

Mean number of clicks counted is shown for each cell in Table 1. For each repetition rate, in either mode, counts increased monotonically with number of clicks presented. Counts were not systematically different from veridical when pairs of clicks were presented at any rate in either mode, nor when any number of clicks were presented at 2/sec monotically or dichotically, or at 4/sec monotically. With further increases in repetition rate and number of presented clicks, counts began to fall short of veridicality, the dropping off being greater for the dichotic than the monotic mode.[3]

For each *S*, for each combination of repetition rate and number of clicks presented,

---

[3] E. A. Holden, Jr., recently has found that educable retardates show more undercounting of dichotically alternating tone bursts than of monotic ones, even at the slow rate of 2.2/sec (personal communication, December 1971).

TABLE 2

Reliability Coefficients (First Half vs. Second Half) of D/M

Repetition rate in clicks/sec.	No. of clicks presented							
	2	3	4	5	6	8	10	20
2	.000	—	.000	.846	—	—	.846	.415
4	.809	.174	.917	.946	.905	.853	.947	.959
6	.584	.940	.941	.816	.554	.765	.861	.811
8	.876	.812	.631	.787	.675	.893	.807	.748

Note.—The tabulated coefficients were calculated by the Spearman-Brown prophecy formula. Blank cells indicate that the original "reliability" coefficients were negative and ignored as without meaning.

the proportion [D/M] was computed of the count in the dichotic mode divided by the count in the monotic mode. The proportion is an inverse index of the amount by which dichotic clicks were undercounted relative to the counts under the monotic condition. The value D/M < 1 indicates relative undercounting of dichotic clicks, D/M = 1 indicates identical counts in the two modes, and D/M > 1 would indicate higher counts in the dichotic than in the monotic mode. The reliability of the index is reflected in Table 2, which gives correlation coefficients, calculated from the Spearman-Brown prophecy formula, between D/Ms derived from the first half of the experiment (5 trials/ cell) and D/Ms from the second half. The variability in the first row of the table (2/sec) likely can be disregarded as reflecting unsystematic factors; the median of the eight coefficients in the row is zero. In the rest of the matrix, only 1 of the 24 coefficients is smaller than r (13) = .441, $p < .05$, and 20 are greater than r (13) = .592, $p < .01$; the median coefficient is .814.

The index D/M is plotted as a function of number of clicks presented in Fig. 1, with repetition rate as the parameter. So far as this figure is concerned, the hypothesis from the motor-response analogy reduces to the specific prediction of a Repetition Rate × Number interaction: The function relating D/M to number of presented clicks should drop off more rapidly as repetition rate increases. This is precisely what occurred. A Rate × Number × *S*s analysis of variance on D/M showed the Rate × Number interaction to be significant, F (21, 294) = 9.60, $p < .001$.

Also significant were the effects of rate, F (3, 42) = 28.72; number, F (7, 98) = 25.61; Rate × *S*s, F(42, 294) = 11.73, ps < .001 and Number × *S*s, F (98, 294) = 1.87, $p < .01$.

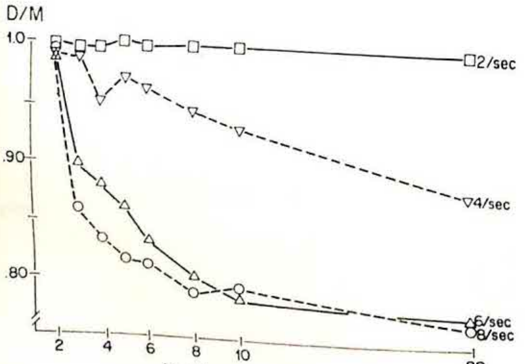


Fig. 1. Count in dichotic mode/count in monotic mode [D/M] as a function of number of clicks in the train, for each of four click-repetition rates.

Discussion

The prediction, that processing of dichotically alternating clicks would be affected negatively by increases in train length and repetition rate, was thus confirmed; and the results are consistent with the notion that repetitive interaural attention shifting can fruitfully be thought of as analogous to a paced, fatigable motor response. This view is in part a restatement of Broadbent's (1954) and Cherry and Taylor's (1954) early suggestions, perhaps prematurely discarded (Huggins, 1964; Moray, 1960; Neisser, 1967), that interaural attention shifts are processes which take time to perform, the time being more or less "dead," i.e., perception being more or less difficult, during the shifts. Interaural shifting of attention has been recently taken up again as an explanatory concept by Moray (1969), and its operation as a time-consuming, performance-limiting process has been emphasized by Treisman (1970, 1971) in interpreting her finding that lists of dichotically alternating digits (4/sec and 6.7/sec) were reported less accurately than lists presented simultaneously to the two ears.

What the motor-response analogy adds, however, is the factor of *fatigability* of the switching mechanism. The most important implication of this addition is that rather than a fixed, constant switching time (Broadbent's [1954] $\frac{1}{2}$ sec., or Cherry & Taylor's [1954] $\frac{1}{6}$ sec.), the time in which a given switch can be performed is seen as dependent upon the number and rate of switches which have preceded it. The concept of fatigability focuses interest on intratrial processes and generates predictions about intratrial performance.

A result of Treisman's (1971) alternating-digit study bears directly on this point. Plotting correctness of recall as a function of serial position in the list, Treisman found that the curve dropped "earlier with fast than with slow presentation, and much more steeply with dichotic than with binaural presentation [p. 162]," a result consonant with the idea proposed here. (A discordant finding, the familiar recency effect manifested in improved recall of the last items of the lists, can probably be attributed to *S*s' ability to tap "echoic memory" [Neisser, 1967] after the trial).

It was suggested earlier that the view of interaural attention shifting as formally similar to a motor process may be more than merely analogical. On a peripheralist view, it could be proposed that the shifting is *underlain* by motor processes, e.g., minimal lateral eye movements, perhaps components of orienting reflexes. (Subjects listening to dichotically alternating stimuli can often be directly observed to voluntarily move their eyes from side to side in synchrony with the onsets of the stimuli, at least with slow stimulus rates.) If this is the case, then the efficiency with which attention shifting can be performed would be determined by the efficiency with which the motor response can be performed. The prediction would then be that the loss in perceptual following of dichotic clicks occurs when they are presented so rapidly that the underlying motor response can no longer keep up with the alternating stimuli.

Kahneman (1970) has made essentially the same point, suggesting that difficulties in following dichotically alternating messages make sense if *S*s are imagined "trying to follow the message with their eyes, [like] spectators at a tabletennis match [p. 126]." From such a view, it is reasonable to expect that ease of following rapidly alternating stimuli would be inversely related to the distance separating the two sound sources, and thus to amplitude of excursion of the orienting movements; and indeed there is evidence that the apparent rate of alternating clicks does increase monotonically toward the apparent rate of monotic clicks as the azimuthal separation between the two sources is decreased (Axelrod & Powazek, 1972).

These considerations need not be limited to audition, of course. They can be invoked to explain the otherwise curious findings of Axelrod, Thompson, and Cohen (1968), who presented pairs of electrical pulses to the fingertips, with a variable interpulse delay (Δt). The *S*s were asked to judge whether the somesthetic "taps" were simultaneous or successive. Mean Δt for successiveness was higher when the pulses were delivered one to each hand (bimanually) than when both were delivered to the same hand. Further, bimanual Δt was higher when the hands were held 114 cm. apart than when they were held 15 cm. apart, and higher still when the forearms were crossed one over the other, this last condition being one in which conflicting and abortive eye orientations would be likely to occur and to be disruptive (Kahneman, 1970).

Support for the motor-process hypothesis of interaural attention shifting can be sought in a number of ways. (*a*) The relative inefficiency of the processing of dichotically alternating inputs should be enhanced if gaze is directed continuously to one side or if lateral-gaze mechanisms are artificially either inactivated (e.g., pharmacologically) or energized out of phase with the auditory stimuli (e.g., with an optokinetic drum). (*b*) The effects of dichotic alternation should be more pronounced in persons with lateral nystagmus than in non-nystagmic persons. (*c*) There are reliable individual differences in D/M (see Table 2); this probably is true as well of the maximum rate at which eye movements can be repetitively driven by moving stripes. The two measures should covary.

REFERENCES

Axelrod, S., & Guzy, L. T. Underestimation of dichotic click rates: Results using methods of absolute estimation and constant stimuli. *Psychonomic Science*, 1968, 12, 133–134.

Axelrod, S., Guzy, L. T., & Diamond, I. T. Perceived rate of monotic and dichotically alternating clicks. *Journal of the Acoustical Society of America*, 1968, **43**, 51–55.

Axelrod, S., & Powazek, M. Dependence of apparent rate of alternating clicks on azimuthal separation between sources. *Psychonomic Science*, 1972, **26**, 217–218.

Axelrod, S., Thompson, L. W., & Cohen, L. D. Effects of senescence on the temporal resolution of somesthetic stimuli presented to one hand or both. *Journal of Gerontology*, 1968, **23**, 191–195.

Broadbent, D. E. The role of auditory localization in attention and memory span. *Journal of Experimental Psychology*, 1954, **47**, 191–196.

Cherry, E. C., & Taylor, W. K. Some further experiments upon the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 1954, **26**, 554–559.

Guzy, L. T., & Axelrod, S. Synchronization of unimanual and bimanual motor responses with monotic and dichotic clicks. *Perception & Psychophysics*, 1971, **9**, 161–164.

Huggins, A. W. F. Distortion of the temporal pattern of speech: Interruption and alternation. *Journal of the Acoustical Society of America*, 1964, **36**, 1055–1064.

Kahneman, D. Remarks on attention control. *Acta Psychologica*, 1970, **33**, 118–131.

Moray, N. Broadbent's filter theory: Postulate H and the problem of switching time. *Quarterly Journal of Experimental Psychology*, 1960, **12**, 214–220.

Moray, N. *Attention: Selective processes in vision and hearing*. London: Hutchinson Educational, 1969.

Neisser, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

Treisman, A. Perception and recall of simultaneous speech stimuli. *Acta Psychologica*, 1970, **33**, 132–148.

Treisman, A. M. Shifting attention between the ears. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 157–167.

(Received December 10, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 295–298

EMOTIONAL IMAGES AS MEDIATORS IN ONE-TRIAL PAIRED-ASSOCIATE LEARNING

EDWARD K. SADALLA[1] AND STANLEY LOFTNESS

University of California, Los Angeles

The *S*s were instructed to construct images with positive, negative, or neutral emotional content for use as mnemonic aids in a paired-associate learning task. Positive and negative images were found to be superior to neutral images in facilitating recall for associated word pairs. No differences were found between the relative utility of positive and negative images.

This study was designed to assess the effectiveness of different types of visual mnemonics on the memory for nouns in a paired-associate (PA) learning paradigm. A number of studies (e.g., Bower, 1970; Ross & Lawrence, 1968) have demonstrated that instructions to use visual images as mnemonic devices facilitate memory for linguistic material (usually lists of nouns, or noun pairs). In such studies, *S*s are instructed to generate a visual image which contains connected visual representations of the words or word pairs presented as stimulus materials. In order to clarify the nature of the learning process involved in such paradigms, a number of recent studies have investigated some of the attributes of instructions to visualize which are effective in promoting performance on memory tasks. Bower (1970) has summarized the principal findings of this research.

One classic prescription for visual mnemonic devices is that the imagined scenes should be "unusual," "bizarre," or "emotionally evocative" (Yates, 1966). Such images were said to be more easily remembered and thus more effective in facilitating memory for other material. Recent investigations concerning the utility of this mnemonic rule, however, thus far have yielded negative results. The *S*s instructed to construct "bizarre" images remember the associated material no better than do *S*s instructed to compose "familiar, regular, sensible associative scenes [Bower, 1970]." No differential effects are obtained when *S*s study pictures of associative scenes which are rated either bizarre or regular by a group of judges (K. Wollen, cited by Bower, 1970). Few data are available concerning image recall as a function of image content. Silverstein (1966) reports that syllables paired with pleasant photographs are no more easily remembered than syllables paired with neutral photographs. The available literature, therefore, does not support the validity of classical intuitions concerning the rules of rhetoric and memory.

The present study investigates the effectiveness of visual mnemonics which vary in "emotional content," rather than "bizarreness." The *S*s were asked to construct associative scenes of either positive, negative, or neutral emotional content in the context of a PA learning task. To the extent that traditional mnemonic instructions are valid, positive and negative images should be more effective as mnemonic aids. The data obtained from the study will be relevant to questions concerning the relationship between "emotional" coding processes and item or image retrievability.

METHOD

Materials and apparatus.—The stimulus materials employed in this study consisted of 54 noun pairs. The nouns were drawn and randomly paired from a set of nouns found to be equal in imagery value, concreteness, and frequency of usage by Paivio, Yuille, and Madigan (1968). The noun pairs were placed on 35-mm. slides and projected on a screen approximately 10 ft. in front of each *S*. A second slide projector was used to project the words, POSITIVE, NEGATIVE, or NEUTRAL, which indicated the experimental condition.

---

[1] Requests for reprints should be sent to Edward K. Sadalla, Department of Psychology, University of California, 405 Hilgard Avenue, Los Angeles, California 90024.

Paired-associate lists.—Six lists of the same 54 noun pairs were constructed for this study. The lists were formed by combining noun pairs with the three words designating the experimental condition such that every noun pair appeared in each experimental condition. This operation yielded three lists which were then ordered such that the sequence of positive, negative, and neutral indicators was random and different for each list. Three additional lists were formed by reversing the stimulus and response members of each noun pair on all three lists. The resulting six lists were thus constructed so that each noun appeared as both stimulus and response term in each experimental condition.

Subjects.—Sixty-six undergraduates at the University of California, Los Angeles, participated in this experiment. All were enrolled in an introductory psychology course, and each received credit toward the satisfaction of a course requirement for participating in the study. The *S*s were assigned to one of the six lists by means of a block-randomization procedure. This procedure insured that each list was included once in each block of 6 *S*s. Two *S*s were omitted from the final data analysis for failure to correctly indicate their answers during the test phase of the study.

Procedure and experimental design.—In the context of a one-trial PA learning paradigm, *S* was presented with one list of 54 noun pairs. Explicit instructions were given concerning the strategies which were to be used while observing and recalling the noun pairs. The experimental manipulations consisted of three different instructional sets which each *S* attempted to apply upon presentation of the noun pairs. The procedure thus involved a mixed-list presentation in which each *S* participated in all three experimental conditions. This procedure corresponds to a single-factor analysis of variance design with repeated measures.

Each *S* was given standard mnemonic instructions (adapted from those described by Bower, in press) to relate the two members of each noun pair through the use of imaginal constructions. In order to test hypotheses concerning the attributes of effective mnemonics, three different types of imaginal codings, positive, negative, and neutral, were requested. The instructions given to *S*s concerning these images were as follows:

> In the NEUTRAL condition form as strong and vivid an image as possible. Always produce an image which relates the two words which appear on the screen. Try not to let the image have any emotional content. In the POSITIVE and NEGATIVE conditions try to form an image *which actually produces a change in your feelings.* In the POSITIVE condition try to form an image which is extremely pleasurable to you, one that makes you happy when you think about it. It should have warm, good, positive emotional content. In the NEGATIVE condition, go all out and deliberately try to scare yourself. The NEGATIVE image should be horrifying and extremely uncomfortable to think about.

All *S*s were given one learning trial and one test trial. Prior to the appearance of each word pair, a word designating the experimental condition was projected for 5 sec. The word pair then appeared for 30 sec. Five minutes after presentation of all the word pairs, *S* was presented with one member of each of the word pairs. The *S* was instructed to first recall the image associated with the word, then to rate the vividness of that image on a 10-point scale, and finally to recall and write out the other member of the word pair associated with the image.

RESULTS AND DISCUSSION

The data analyzed were the mean number of errors per *S*, the imagery ratings of recalled images, and the number of complete failures of image recalls. An error of word recall was defined as any word which did not exactly correspond to the word in the original stimulus list. In many cases, *S*s retrieved words which were highly similar to the original item, for example, the recall of the word "beach" instead of the correct word, "shore." In such cases, *S*s apparently retrieved the original imaginal encoding but failed to associate it with the correct verbal term. An analysis was performed on the obtained data with these "near misses" counted as correct responses. That analysis did not yield different results from those described below.

The data obtained in this study are depicted in Table 1. The mean number of errors per *S* was subjected to an analysis of variance for correlated means. That analysis yielded a main effect for image type, $F\ (2,\ 126) = 9.02$, $p < .001$, indicating that instructions to construct emotion-laden imaginative associations influenced performance on the PA task. Instructions to construct positive and negative images increased recall performance

TABLE 1

MEAN ERROR SCORES, MEAN NUMBER OF FAILURES OF IMAGE RECALL, AND MEAN VIVIDNESS RATINGS FOR RECALLED IMAGES PER *S*

Measure	Experimental cond.		
	Positive	Negative	Neutral
Errors	5.84	5.78	7.09
Failures of image recall	3.04	3.72	4.61
Vividness ratings	7.15	7.32	6.37

over instructions to construct neutral images. No differences in recall were found to result from the difference between positive and negative imaginal associations. These findings are consonant with the classical instructions to use emotionally evocative, unusual, or bizarre images rather than ordinary scenes as mnemonics for encoding to-be-remembered material (Yates, 1966).

The number of instances in which no image was reported also varied between conditions. Neutral images were associated with more complete failures of recall than were either positive or negative images, $F\ (2,\ 126) = 13.9,\ p < .001$. This difference in retrievability accounts for the major portion of the variation in errors between experimental conditions. Sixty-four percent of the total number of errors committed occurred on trials when no image was reported. A Pearson rank-order correlation, computed between imagery rating and number of errors, yielded a rho of .82. It should be noted, however, that this relationship might be spuriously high as a result of *S*s' confusing the imagery rating with information pertaining to the difficulty of word recall.

The difference between the results found in this study and those reported in previous investigations of mnemonic devices may result from differences in experimental design and procedure. The most obvious difference involved the use of instructions to create "emotional" rather than "bizarre" imaginal associations. Bower (in press) points out that "bizarreness" is a poorly defined term which probably elicits diverse behaviors on the part of an *S* population. Postexperimental interviews suggested that the instructions utilized in this investigation were clearly understood and easily carried out by most of the participating *S*s. Other procedural differences included the use, in this study, of longer encoding times combined with a design in which each *S* participated in all mnemonic conditions. Within-condition variance was minimized through the use of nouns which were equated for imagery evoking potential.

The results obtained in this study indicate that given particular instructional and procedural conditions, *S*s are significantly better at recalling PA items which have been related through emotionally evocative imaginal associations. This conclusion necessarily presumes that the manipulated instructions affected the representation of PA items in memory.

An alternative interpretation involves the possibility that instructions to construct emotional encodings exert their effect on rehearsal strategies, perhaps by compelling differential attention to "emotion" tagged items. Within-*S* designs involving instructional manipulations may produce a tendency for instructional sets appropriate to one experimental condition to carry over and influence other experimental conditions. The extent of systematic cross-condition influence was mitigated by the use of three different randomly generated sequences of instructional sets. The possibility remains, however, that the range of convenience of this mnemonic rule is limited to contexts involving either a within-*S* or mixed-list design. Exploration of these procedural variations might clarify questions concerning the process by which emotion related instructional sets exert their influence on memory.

Instructions to image are presumed to initiate a sequential coding process, which requires some temporal interval for completion (Bugelski, Kidd, & Segmen, 1968). At the time of recall, the stimulus word is assumed to lead to the redintegration of a short autobiographical episode, during which *S* remembers the cognitive elaborations he generated while studying the noun pair (Bower, in press). The data from the present study indicate that instructions to create "emotional" associations lead to the construction of a sequence of cognitive elaborations which are later recalled more easily and with greater rated clarity. The ease and rated clarity of imagery recall are, in turn, strongly related to the probability that *S* will correctly recall the response member of the original noun pair.

REFERENCES

Bower, G. H. Analysis of a mnemonic device. *American Scientist*, 1970, **70**, 496–510.

Bower, G. H. Mental imagery and associative learning. In L. Gregg (Ed.), *Cognition in learning and memory.* New York: Wiley, in press.

Bugelski, B. R., Kidd, E., & Segmen, J. Image as mediator in one-trial paired-associate learning. *Journal of Experimental Psychology*, 1968, **76**, 69–73.

Paivio, A., Yuille, J. C., & Madigan, S. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 1968, **76** (1, Pt. 2.)

Ross, J., & Lawrence, K. A. Some observations on memory artifice. *Psychonomic Science*, 1968, **13**, 107–108.

Silverstein, A. Acquired word value, verbal learning and retention: Replication and extension. *Psychological Reports*, 1966, **18**, 357–358.

Yates, F. *The art of memory.* Chicago: University of Chicago Press, 1966.

(Received December 16, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 299–307

TRACKING OF SERIAL PATTERNS [1]

FRANK RESTLE [2] AND BILLY L. BURNSIDE

Indiana University

College students saw serial patterns of six lights and attempted to track them by pressing the appropriate button as the light came on, at a rate of one light per .7 sec. The main indicator of difficulty was an error or omission immediately after a correct response (a C-E transition) indicating that *S* got off the track at that point. The data showed that (*a*) the beginning and sometimes the end of a subunit (run, trill, or repetition) is difficult, (*b*) organized arrangement of subunits leads to far better performance than disarrangement of the subunits, (*c*) tendencies to use runs versus trills at one point in a sequence depend upon the frequency of such subunits elsewhere in the sequence, (*d*) error patterns faithfully mirror a hierarchical organization built into the sequence, and (*e*) if a single element of a sequence deviates from a hierarchical tree, errors are found concentrated in the branch of the tree affected. The results span serial pattern learning, perception of sequential patterns, and coordinated motor skill.

When a college student is required to perform a serial pattern by anticipating each element, he develops a complex cognitive structure, hierarchically organized (Restle, 1970; Restle & Brown, 1970a). This conclusion is based on experiments in which *S* predicts which of six lights in a panel will come on next, and indicates each prediction by pushing a corresponding button. It was found that the overall difficulty of learning a serial pattern, the relative difficulty of locations in the patterns, and the particular errors made, all depend upon the structure of the pattern being learned (Restle & Brown, 1970a, 1970b, 1970c). Furthermore, the data reveal the existence of complicated hierarchical hypotheses (Restle & Brown, 1970a).

There are several ways that the serial anticipation experiment may be criticized as being an atypical example of serial learning. First, the elements of the serial pattern are presented quite slowly (about 4 sec. per individual item) so that the pattern does not seem to stand out "perceptually." In this, serial pattern learning differs from musical patterns, which occur much more rapidly and are perceived rather than being apprehended in a more deliberate and conscious manner. Second, in serial anticipation *S* is required to make active anticipations even before he has learned the pattern, and this anticipation may interfere with the natural process of apprehension, as well as imposing a structure that otherwise might not occur in the process. Finally, in serial anticipation learning, most of the errors are responses made before *S* has found the pattern, and therefore may be thought to be irrelevant to the psychological structure of the pattern.

These three criticisms do not totally invalidate the results of the earlier experimental program, but it seems useful to study the cognition of serial patterns by a second method that escapes these difficulties and to determine if the same conclusions can be drawn from this alternative method.

The question asked, then, is whether the hierarchical structures found in earlier studies are attributed to deliberate training and the development of premature hypotheses or if they are characteristic of serial pattern learning in general. Suppose, for example, that the structures in the previous study were found to depend upon verbal rehearsal and that the patterning found in data arose from the use of certain conventional linguistic forms for the encoding and storing of information. If this were true, then a faster task, especially if it served to inhibit verbal encoding and rehearsal, would eliminate such patterning.

---

[1] This research was supported by United States Public Health Service Grant PHS MH-16817.

[2] Requests for reprints should be sent to Frank Restle, Department of Psychology, Indiana University, Bloomington, Indiana 47401.

The alternative hypothesis is that pattern learning necessarily involves the eduction of relationships from the elements of the pattern and that these relationships are themselves interrelated in a hierarchical structure. This position states, in effect, that the cognition of a pattern resides in the perception of the pattern itself, rather than outside in a linguistic description.

To distinguish these two possibilities, we had *S*s see patterns that were presented very fast, in a circumstance in which it was unlikely that verbal habit systems had a controlling influence. Garner and Royer (1966) studied binary sequences by presenting the two alternatives rapidly and having *S*s make responses concurrently with the events, a process that may be called "tracking." Tracking, using only about 1 sec. per event, is so rapid that *S* is likely not to be able to control it by verbal means, so that any structure found in tracking experiments cannot plausibly be attributed to verbal descriptions or formulated hypotheses. Whatever structure is found in tracking data can be attributed to cognitive processes more nearly intrinsic to the perception of serial patterns, not verbal description and elaboration.

Whereas in anticipation learning the successive responses are mainly independent, in tracking the various responses are integrated into a continuous flow. Once this flow is interrupted in tracking, *S* may then miss the next several responses until he can reestablish his "rhythm," and will often fail at an easy location that follows a difficult one. Mere frequency of failure is therefore not a valid indication of difficulty at that specific location, and we instead count "correct-error" (C-E) pairs, errors that immediately follow a correct response. It is also possible, in tracking, to determine where it is that *S*s get back onto the track, i.e., to investigate "error-correct" (E-C) pairs.

Because tracking is fast, *S*s can respond late. The response intervals surround the events themselves, and there is no separation between response intervals, so that it is quite possible for *S* to respond late, and thereby place the response to Location N in the response interval reserved for Location $N + 1$. The way our apparatus was programmed, if *S* then makes another response in that same interval, catching up with the sequence, his second response would not be recorded. In most sequences, the result is that *S* appears to have made two errors in succession, a blank (failure to respond) followed immediately by the response that would have been correct at the previous location, and this usually followed by a correct response. The exception is a sequence in which the same response is correct twice in a row, for then if *S* is late in responding to the first, he inadvertently will be correct at the next location.

Hypotheses.—Our investigation of the cognitive structure of serial tracking dealt with four basic questions. First, are such sequences divided into subparts, and is it true that runs, trills, and repetitions form such subunits? In this initial study we sought to learn what particular patterns of data correspond to such subunits, if they are different from what is seen in anticipation learning.

Second, if it is at least true that short runs form subparts, can it be shown that tracking behavior is influenced by regular (higher level) organization of such subparts, as contrasted with an irregular arrangement?

Third, if sequences are organized into parts, are the parts fixed or are they labile? Can the choice between alternative possible organizations be influenced by other parts of the sequence?

Finally, we studied tracking of some sequences organized by hierarchical trees. In one experiment we investigated several different versions of what is essentially the same hierarchical structure, asking if (as in anticipation learning) locations at the apex of the hierarchical tree were most difficult, and those at its base easiest. In another study, we investigated the effects of "faults" or exceptions in the sequence, that is, sequences in which all elements fit a single hierarchical pattern except for one location.

GENERAL METHOD

Subjects.—Undergraduate students from introductory psychology courses at Indiana University served as *S*s in these experiments, in partial fulfillment of course requirements. The *S*s were tested in squads of 4 or fewer, and all *S*s in a squad were tested on the same sequences presented in the same order. Experiments I, III, and IV used the same group of 65 *S*s, Exp. II used 39 and Exp. V used 47 others, all drawn from the same pool, for a total of 151 *S*s.

Apparatus.—The experiment was conducted in a small room containing four individual *S* booths, so that each *S* was isolated from his neighbors by a plywood partition. On a slanting panel in front of each *S* was a row of six small white event lights, with a push button directly below each of them. A tone generator was used to warn *S*s of the start of each sequence. The event lights and tone generator were controlled from another room by an IBM 1800 process-control computer, and responses were sensed and recorded by that same device.

Procedure.—The *S*s were instructed that they would hear a warning tone and then the event lights would flash in a repeating pattern they were to track by pressing the button under each light as it flashed. It was emphasized that the button-press responses must occur concurrently with the light flashes. The *S*s were told that they could watch the lights for as long as they wanted before starting to respond, so that they had an opportunity to learn each pattern before responding to it.

In each experiment, *S*s were first presented with several repetitions of a practice sequence, followed by several repetitions of one or more test sequences. The *E* was present in the experimental room during presentation of the practice sequences, but not during presentation of the test sequences. All sequences were visually presented to *S*s by flashes of the six event lights, with each flash having a duration of .1 sec., and with .6 sec. between flashes. There was no break between successive repetitions of a single sequence, but *S*s were given a 20-sec. rest between presentations of different sequences. Responses to light flashes were recorded during a period from .3 sec. prior to each flash until .3 sec. after the flash (for a total of .7 sec.), so that both anticipatory and delayed responses were recorded.

EXPERIMENT I: SUBUNITS

The purpose of Exp. I was to determine whether *S*s would form sequences in a tracking experiment as they do in serial anticipation learning of patterns. The sequential patterns used were composed of runs, trills, and blocks of repetitions of the same event, arranged in three different orders.

Method.—The *S*s first saw 15 repetitions of a practice sequence (12342323), which contains both a run and a trill, and later[3] saw 25 repetitions of one of the three sequences shown in Table 1. These sequences differ only in the order of the subunits and in the makeup of one short subunit, which was either a run (543) or a "random" subsequence (164).

TABLE 1

SEQUENCES USED IN EXPERIMENT I (SUBUNITS)

Location	Sequence		
	a	b	c
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	6	6	1
6	6	6	6
7	6	6	4
8	6	6	2
9	6	6	3
10	2	2	2
11	3	3	3
12	2	2	6
13	3	3	6
14	5	1	6
15	4	6	6
16	3	4	6

Note.—n = 17, 24, and 24 for Sequences a, b, and c, respectively.

Results.—The *S*s watched the sequences a mean of 3.55 times through, letting an average of 56.8 lights flash before making their first tracking response. Although Sequence a appears the most regular, *S*s waited longest (mean of 3.83 times through) before responding to it. After *S*s started responding they were .73 correct, and on the last five trials they made an average of .78 correct, indicating that most of the learning took place while *S*s were watching and before their first response. Approximately .90 of all errors were anticipatory responses, late responses, or failures to respond, which suggests that *S*s may have had a fairly clear idea of the sequence, even better than .73 or .78 correct might suggest.

---

[3] The 65 *S*s who participated in Exp. I, III, and IV saw one of the sequences used in each of these experiments, all in one session. Sequences were assigned to *S* by squads by means of a Latin square design so that, e.g., a squad might have seen Sequence a from Exp. I, Sequence b from Exp. II, and Sequence c from Exp. IV. No *S* saw more than one sequence from the same experiment, or more than one sequence represented by the same letter (a, b, or c). The three sequences presented to a squad of *S*s were presented in a random order.

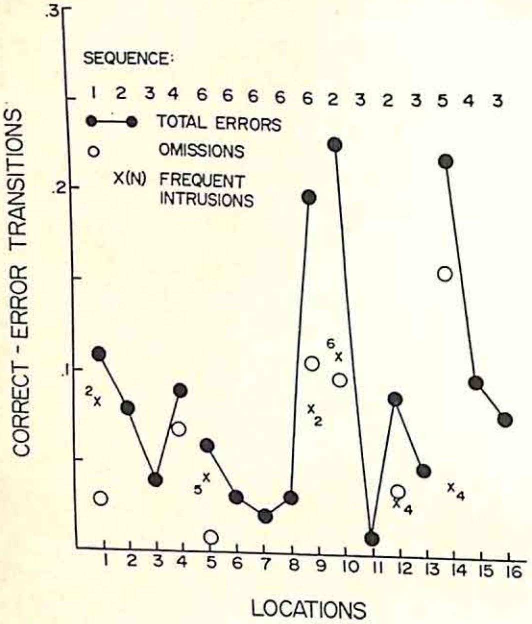


FIG. 1. Correct-Error transition proportions for each location in Sequence a. (Omissions and high-frequency errors are indicated along with total errors.)

The proportion most clearly showing the difficulty of a given location is the probability that an error is made at that location, given that a correct response was made at the immediately preceding location. We call this the C-E probability, signifying a correct response followed by an error, and say that *S* is going off the track with this probability.

If tracking is like anticipation learning, *S*s will make most errors at the beginning of a run (and perhaps anticipation errors at its end) and may make run-overextension errors after the end of the run. All of these effects are found in the data of sequence shown in Fig. 1. Notice that at Locations 1–4 and 14–16, *S* is learning a run. Error frequencies decrease during the run, except at Location 4, the end of the long run. Furthermore, at Location 5, the high-frequency error is response "5," overextension of the run, and at Location 1 (which follows Location 16) the most frequent error is "2," again the overextension of the run.

Our studies of anticipation learning have not isolated a definite response pattern for trills, except that *S*s frequently show some run overextension errors during trills, as are shown in Locations 12 and 14.

Finally, the block of consecutive 6's shows a decreasing frequency of errors until near the end. The last 6 frequently is missed as *S*s anticipate the next event, and if *S* does not miss that location he then frequently makes an extra repeated response. From this, one may conclude that a block of consecutive identical events behaves like a run, except that with a run *S* can tell when to terminate by noticing that the run ends on a particular event, whereas a block of identical events has no such "marker" for its termination. Essentially the same results were found in Sequences b and c, with high C-E probabilities being found at all locations in the "random" subunits. So, in summary, the C-E proportions, corresponding to the probability that *S* gets off the track at each location of the sequence, are very like the overall errors found in anticipation learning, and this supports the hypothesis that tracking depends upon a cognitive structure essentially like that of anticipation learning.

The probability of getting back on the track, E-C proportions, turned out to be

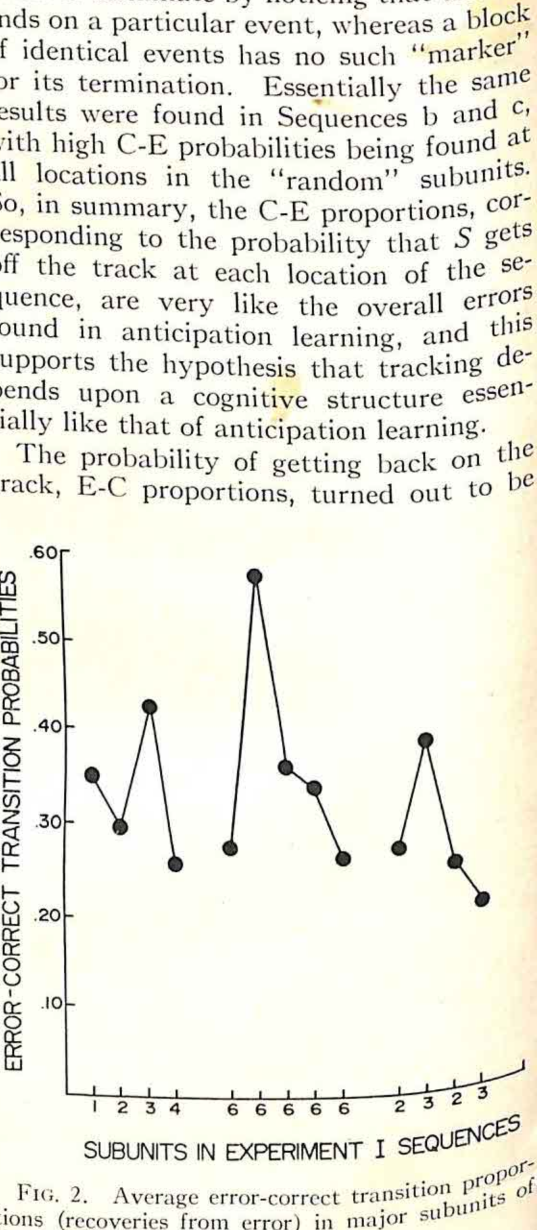


FIG. 2. Average error-correct transition proportions (recoveries from error) in major subunits of the three sequences in Exp. I.

less informative. The run of 4, the block of 5 identical events, and the trill, were abstracted from all three sequences, and the E-C proportions averaged. The results, shown in Fig. 2, indicate that *S*s do not most often rejoin the sequence at the beginning of a subunit, but instead at the second or third location within it. This is particularly clear in the case of the block of identical events, and least clear in the case of a run.

EXPERIMENT II: ARRANGEMENT OF SUBUNITS

While Exp. I showed how subunits affect tracking, it did not show that any higher-level organization is effective. Does tracking depend only on the subunits used or also on the higher order patterns into which subunits may be fit? The method of Exp. II was to contrast a highly organized pattern with a sequence produced by disarranging the subunits, without anywhere disturbing the integrity of any subunit.

Method.—The 39 *S*s were first presented with 25 repetitions of the practice sequence (1 2 3 4 2 3 2 3 6 5 4 3 5 4 5 4). This was followed by 20 repetitions of the test sequence, which either could be organized or not. The organized sequence, grouped by parentheses, is as follows: (1 2 3) (2 3 4) (3 4 5) (4 5 6) (6 5 4) (5 4 3) (4 3 2) (3 2 1). The unorganized sequence was made up of the following disarrangement of the same subunits: (4 3 2) (4 5 6) (1 2 3) (6 5 4) (2 3 4) (3 2 1) (3 4 5) (5 4 3). The organized sequence was given to 20 *S*s, the unorganized to 19.

Results.—Higher order organization has even a more dramatic effect in tracking than was found in serial anticipation. The *S*s waited through 1.96 trials of the organized sequence before responding, but 5.93 trials of the unorganized sequence, and a *t* test showed this difference to be significant, $t\ (37) = 3.8$, $p < .001$. This long waiting period did not merely reflect caution, for whereas *S*s were .63 correct on the organized sequence after they began responding, they were only .46 correct on the unorganized sequence. A *t* test on differences in number of correct responses between the two groups was significant, $t\ (37) = 4.27$, $p < .001$.

TABLE 2

SEQUENCES USED IN EXPERIMENT III (AMBIGUOUS SEQUENCES)

Location	Sequence		
	a	b	c
1	2	2	2
2	1	1	1
3	2	2	2
4	3	3	3
5	4	4	4
6	3	3	3
7	4	4	4
8	5	5	5
9	6	4	6
10	5	5	5
11	4	4	4
12	3	3	3
13	4	4	2
14	3	3	3

Note.—n = 30, 16, and 19 for Sequence a, b, and c, respectively.

The larger number of C-E transitions found on the unorganized sequence was concentrated at the beginnings of run subunits (Locations 1, 4, 7, etc.). The average proportion of C-E transitions at these locations in the organized sequence was .24, while the corresponding proportion in the unorganized sequence was .43.

EXPERIMENT III: AMBIGUOUS SEQUENCES

Experiment III raises the question whether the organizing structures within subunits in tracking are fixed innate tendencies with fixed strength or whether they can be modified to fit details of the sequence. The experiment centers around an "ambiguous" sequence that can be organized into either trills or runs.

Method.—The *S*s were the same 65 undergraduates used in Exp. I assigned to either Sequence a, b, or c. The sequences and number of *S*s assigned to each are shown in Table 2. Notice that Pattern a can be organized either as a trill or as a run; Pattern b favors organization by trills, and Pattern *c* favors organization by runs. Also notice that the three patterns are very similar, differing only at Locations 9 and 13.

Results.—Particular interest centers on those locations, common to the three patterns, at which trill or run overextension errors can be made. If these organizing tendencies are labile, we should expect more trill responses in Pattern b, more run responses in Pattern c, and an intermediate frequency in Pattern a.

At Location 6 of all three sequences it is possible to make response "5," overextending the preceding run (1 2 3 4). The frequency of this response, as a C-E pattern, is .16 in Pattern c in which such run errors should be most frequent, .10 in Pattern b in which trill errors are emphasized, and .09 in Pattern a, neutral. Thus, overemphasis on runs in Pattern c does produce an excess of such responses.

At Location 8 in all three sequences it is possible to overextend a trill sequence, making response "3" after the trill (3 4 3 4). The frequency of this response as a C-E pattern in Pattern b in which such trill tendencies were emphasized is .08, whereas in Patterns a and c such responses are rare (.01 each).

The above results indicate that even in a segment of the pattern that is identical among the three patterns, differences in response frequency can be found and the differences are in the direction expected if the strength of run and trill tendencies is adjusted to the overall structure of the sequence.

EXPERIMENT IV: HIERARCHICAL TREES

A major finding from the studies of serial anticipation learning of patterns was the regular distribution of errors made when the sequence could be described by a binary hierarchical tree. In such a sequence, each light is one of a pair, each pair is one of a pair of pairs, each such unit is one-half of a larger unit, etc. The Sequence (1 2 1 2 6 5 6 5) would come about by starting with Event 1, then transposing (T) it to obtain T(1) = (1 2), then repeating (R) this giving R(T(1)) = (1 2 1 2) and then doing a mirror image (M) of that, yielding M(R(T(1))) = (1 2 1 2 6 5 6 5).

Method.—Three sequences of Length 16 were constructed using the operations of transposing, repeating, and mirror imaging, but in different orders. They are shown in Table 3. Pattern a is described as T(M(R(T(1)))), Pattern b as T(M(T(R(1)))), and Pattern c as T(R(M(T(1)))).

Results.—If tracking follows the kind of complex cognitive structuring found with anticipatory learning, then the most difficult location will be that at which *S* must use his highest level decision. The highest level decision controls the transition from the first to the second half of the sequence, and a failure to respond at Location 9. The next highest level decision controls quarters of the sequence, and faults would appear at Locations 5 and 13. The next highest level would produce difficulties at Locations 3, 7, 11, and 15, and the lowest level would produce difficulties at the even-numbered locations.

TABLE 3

SEQUENCES USED IN EXPERIMENT IV (HIERARCHICAL TREES)

Location	Sequence		
	a	b	c
1	1	1	1
2	2	1	2
3	1	2	6
4	2	2	5
5	6	6	1
6	5	6	2
7	6	5	6
8	5	5	5
9	2	2	2
10	3	2	3
11	2	3	5
12	3	3	4
13	5	5	2
14	4	5	3
15	5	4	5
16	4	4	4

Note.—n = 18, 25, and 22 for Sequences a, b, and c, respectively.

In fact, C-E (and correct-zero) transition probabilities for Sequences a and c closely follow this predicted pattern, as is shown in Fig. 3. The higher the location is in the structural tree, the more errors there are, and this regular trend is found both in omissions and in overt errors.

The profile of errors on these sequences is given in Fig. 4, which shows how consistently the number of C-E errors follows the hierarchical structure of the sequence. This pattern of data is as clear as that found in anticipation learning.

The results in Pattern b are not in complete agreement with those above, but for an instructive reason. In Pattern b each individual event is repeated immediately, (1 1 6 6...). In tracking it is quite common for *S* to respond a little late, in which case, normally, he would be scored as an omission on the interval in question (having failed to respond) and a "perseveration" error at the next location. However, when the same event follows itself, a response to the first that is late will be scored as correct in the next interval. This results in an artificially high accuracy in the second of two repetitions. Furthermore, it appears that this encourages *S*s to make more perseveration errors than would otherwise occur.

Despite this discrepancy it is clear from Exp. IV that tracking of a hierarchically organized serial pattern produces errors that are also hierarchically organized, and the result is very similar to what was found by anticipation learning. Therefore, we may conclude that the cognitive structure used to control tracking not only divides the sequence into parts, takes advantage of regular arrangement of the parts, and is modifiable to the sequence; but in addition, the tracking-control system can develop and use the complex hierarchical trees that develop with slower anticipation learning.

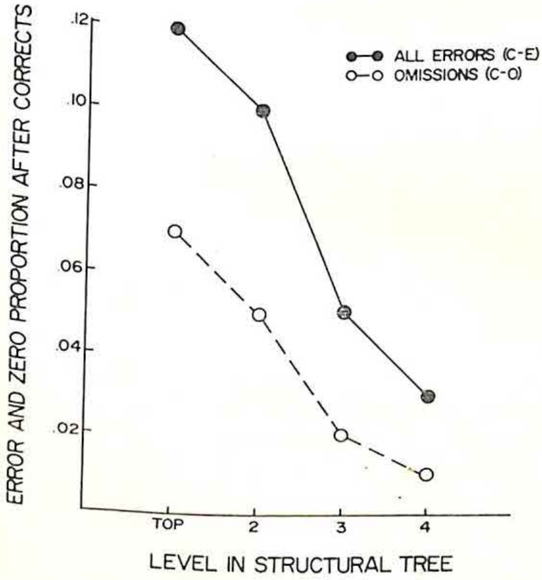


FIG. 3. Correct-Error transition proportions and correct-omission transitions as a function of level in structural tree (Sequences a and c of Exp. IV).

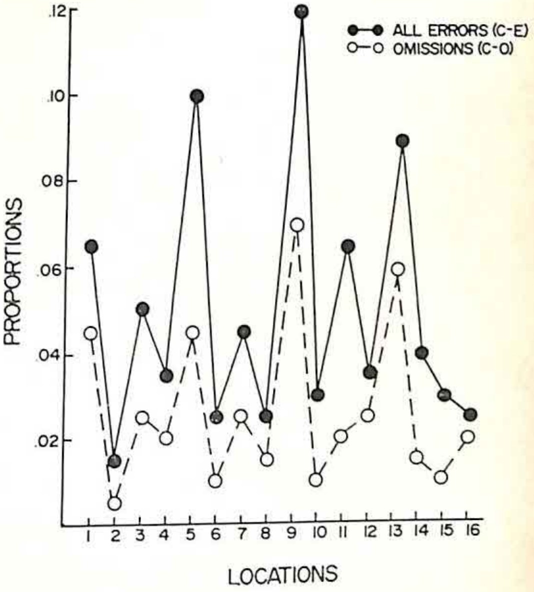


FIG. 4. Profile of correct-error and correct-omission transition proportions as a function of location in Sequences a and c of Exp. IV.

EXPERIMENT V: HIERARCHICAL STRUCTURES WITH FAULTS

Consider a sequence having a completely hierarchical structure except for one location. This is, in effect, a perfectly organized sequence except for a single fault. In anticipation experiments, *S*s make a disproportionate number of errors at the particular location containing the irregularity, and furthermore tend to intrude the event that would have been correct at that location if the sequence had had a perfect hierarchical structure. This is our most conclusive and direct evidence of the use of an hierarchical cognitive structure.

Method.—Using the procedures of the previous experiment, 47 new *S*s each tracked one of the three sequences shown below:

Sequence d 1 2 1 2 2 3 2 3 6 5 6 5 5 4 5 4
Sequence e 1 2 1 2 2 3 2 2 6 5 6 5 5 4 5 4
Sequence f 1 2 1 2 3 3 2 3 6 5 6 5 5 4 5 4

Pattern d is a regular hierarchical tree, and Patterns e and f have single discrepancies. Pattern e has a repetition leading to Location 8 where the hierarchical pattern would require a transposition. Pattern f has a discrepancy at Location 5, where the transposition is applied to the past event instead of to the whole first subsegment (1 2 1 2).

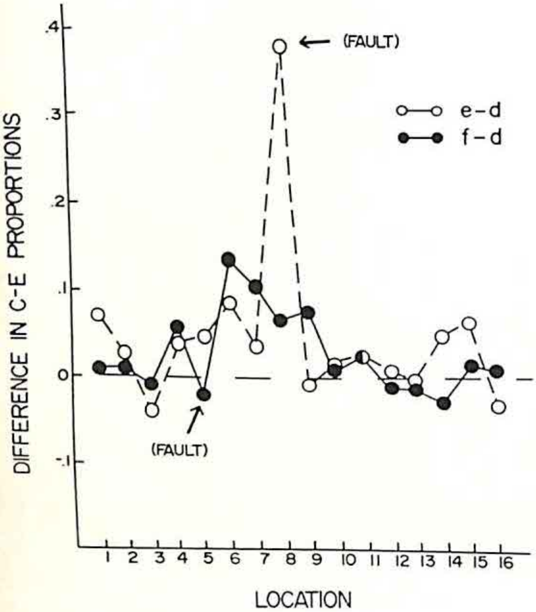


FIG. 5. Effects of faults in regular hierarchical sequences. (Differences in frequency of correct-error transitions between experimental (faulty) and control (perfectly hierarchical) sequences. Sequence e had its faults at a low structural level, Sequence f at a high level.)

Results.—If these *S*s are using hierarchical structures, then this should be reflected in the error pattern of Pattern d, which is regular. Pattern e should differ from Pattern d at a single location, Location 8. The effect is local since the discrepancy is at the lowest branch of the tree, therefore not disturbing any other part of the tree. The effect to be expected of Pattern f is more complicated since the entire subunit, Location 5–8, may be disrupted. At Location 5, where the discrepancy in the pattern is located, *S* will not have a particularly strong cognitive expectation until he has really mastered the sequence, for this is a relatively difficulty location (second level of the hierarchical tree). Therefore, errors may not pile up at that particular location. However, Locations 6, 7, and 8 are branches of the bough that was disrupted by the discrepancy, and might show more than the average number of errors.

Pattern d was a regular hierarchical tree and showed essentially the same profile of C-E errors as are given in Fig. 4. Interest in this experiment centers upon the differences between d and the faulty Patterns e and f. Therefore, we consider performance on Pattern d to be a "control" and subtract the number of C-E errors at each location of Pattern d from the number of C-E errors at the corresponding location of Pattern e or f. These excess errors attributable to faults in Patterns e and f are shown in Fig. 5.

The only large difference between Pattern d and Pattern e occurred at Location 8, where an abnormally high frequency of C-E errors were made on Pattern e. Furthermore, a very high proportion of these errors, .18 of .39 in fact, were response "3," which is the event that would have been there if the pattern had been perfectly regular. The results of tracking Pattern e, therefore, conform perfectly to what should be expected if *S*s use hierarchical cognitive structures to track patterns like this.

As expected, Pattern f did not produce a particularly high frequency of errors at the location containing the fault, but did produce rather large discrepancies between Pattern d and Pattern f at locations throughout the subunit containing the fault (Locations 6, 7, and 8). No large proportion of any particular error was found at these locations.

The results of Exp. V clearly show that *S*s are using hierarchically organized cognitive structures in tracking these sequences. When a low-order event is changed, *S*s make a great many errors at that location, and the errors they make consist of filling in the response that the hierarchical structure generates. When a high-order event is changed, errors are increased at all locations in the subunit containing that changed location.

It would seem natural for many psychologists to agree that *S*s "organize" the serial patterns, but to apply a generalized "gestalt" theory. In this context, Exp. V appears to be a study in "closure," the tendency to regularize the perception of an irregular pattern. However, if this level of theory were sufficient, we should expect that Patterns e and f, each of which is produced by changing a single element of the same regular pattern, would produce equivalent experimental results. The concept of hierarchical tree structure, on the contrary, points out that the faults in Patterns e and f are at structurally different points in the pattern and should have quite different consequences on behavior. The results of the experiment clearly indicate that the two kinds of faults in the pattern result in quite different kinds of error on the part of *S*. This evidence is not enough to prove the detailed theory of hierarchical trees, but it is sufficient to prove that a simple, one-level idea of pattern or organization of the sequence is not adequate.

DISCUSSION

The experiments reported in this article constitute a replication of the main results reported by Restle and Brown (1970a, 1970b, 1970c) and Restle (1970) with regard to serial pattern learning by anticipation. All of the main indicators of cognitive structure found in serial pattern learning have now also been shown to occur in serial tracking. This strongly suggests that the findings are not the result of slow, conscious, verbal encoding and manipulation of the events in the serial pattern, but instead result from a rapid process by which serial information is organized.

Perceiving and learning serial patterns.—The results mean, first, that the processes of serial pattern organization are closely related to pattern perception, and consequently to the perception of music, poetry, dance, and other organized serial patterns. The earlier efforts to relate a serial pattern learning to music (Restle, 1970; Restle & Brown, 1970a) are therefore supported.

Tracking as an integrated skill.—The second main implication of these results has to do with the integration of a skilled task. The task of tracking the lights in this experiment, to *S*, certainly seems demanding and requires a fairly high degree of "coordination" of motions, at least resembling coordinated skills.

The findings suggest three hypotheses regarding the nature of such a coordinated skill. First, that perception of the required pattern is important, since our *S*s learn more by observing the pattern than trying to perform it.

Second, such skills are controlled by a rule structure, dividing the task into subunits each of which is governed by a simple rule system. Subunits, in turn, may be related to one another by higher order rules, and learning proceeds far more rapidly and to a higher asymptote when a coherent overall organization exists, than when the same subunits are disarranged. The subunits of a task are not at all arbitrary, but instead depend upon the inner structure of the task itself, that is, upon the degree to which the elements of the task are objectively arranged according to a rule system.

The third main hypothesis is that such rules, for the generation of a complex sequence, can be organized into a hierarchical tree, and that the learning and performance of a task depends upon the nature of that tree structure. The most difficult points or transitions in the task will correspond to the major divisions of the tree. Furthermore, if the actual task deviates from a coherent tree in a specific way, *S*s are likely to concentrate their errors at this discrepancy and all other parts of the sequence that have an intimate structural relationship to the discrepancy.

The central theoretical importance of the structural hypotheses put forward in this work is illustrated by the fact that the same concepts are used, with equal appropriateness, in the description of perception, learning, and motor skill. Such ideas, that leap across the chapter divisions of conventional psychological thought, may eventually serve as the means for integrating a general psychological theory of behavior.

REFERENCES

GARNER, W. R., & ROYER, F. L. Response uncertainty and perceptual difficulty of auditory temporal patterns. *Perception & Psychophysics*, 1966, **1**, 41–47.

RESTLE, F. Theory of serial pattern learning: Structural trees. *Psychological Review*, 1970, **77**, 481–495.

RESTLE, F., & BROWN, E. R. Organization of serial pattern learning. In G. H. Bower & J. T. Spence (Eds.), *Psychology of learning and motivation.* New York: Academic Press, 1970. (a)

RESTLE, F., & BROWN, E. R. Serial pattern learning. *Journal of Experimental Psychology*, 1970, **83**, 120–125. (b)

RESTLE, F., & BROWN, E. R. Serial pattern learning: Pretraining of runs and trills. *Psychonomic Science*, 1970, **19**, 321–322. (c)

(Received December 17, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 308–316

DETERMINANTS OF INDUCED AMNESIA IN SHORT-TERM MEMORY[1]

DOUGLAS K. DETTERMAN[2]
University of Dayton

NORMAN R. ELLIS
University of Alabama

In two experiments, *S*s were required to free recall lists of line drawings of common objects. Some of the lists contained a critical item, a photograph of nudes, in place of the middle item. In addition, instructions regarding rehearsal, rate of presentation of list items, and exposure duration of the critical item were varied. The critical item produced a von Restorff effect accompanied by substantial anterograde and retrograde induced amnesia. Instructions minimizing rehearsal had no effect on retrograde amnesia, but increased anterograde amnesia. Faster rates of presentation increased only anterograde amnesia while longer exposure durations of the critical item increased only retrograde amnesia. These results were interpreted as suggesting separate causal factors for retrograde and anterograde amnesia. Retrograde amnesia seemed best explained in terms of retrieval failure and anterograde amnesia in terms of encoding failure.

In some memory and learning tasks, when an unusual or outstanding critical event occurs in an otherwise homogeneous series of events, memory for the outstanding event is facilitated; i.e., a von Restorff effect occurs (Wallace, 1965). In addition, memory for events before and after the critical event is often markedly impaired. This impairment is demonstrated by comparing retention of lists with critical items to that of those not having a critical item. This decrement in memory, referred to as induced amnesia, has been suggested as an experimental analog to clinical amnesia.

Several investigators have reported decrements in retention of items surrounding a critical item (Jenkins & Postman, 1948; Tatuno, 1961), but Tulving (1969) was the first to explicitly demonstrate induced retrograde amnesia. Before participating in a free-recall task, *S*s were instructed to remember a high-priority critical event (the name of a famous person) if it occurred in any of the lists of common nouns they were presented. Though a retrograde amnesic effect was obtained, there was only a slight indication of an anterograde effect. Saufley and Winograd (1970) have replicated these results without giving *S*s special instructions to remember the critical item.

Ellis, Detterman, Runcie, McCarver, and Craig (1971) demonstrated larger amnesic effects using a somewhat different procedure. Instead of having *S*s remember high-priority events, they used photographs of nudes as critical items. Their first experiment employed a free-recall task and demonstrated both anterograde and retrograde effects. Only anterograde amnesia was found with a recognition task in the second experiment.

Theoretical Explanations

In light of available data, several hypotheses might reasonably account for the amnesic effect.

Consolidation.—Tulving's (1969) hypothesis that the retrograde amnesic effect represented a disruption in consolidation processes is consistent with two findings from his experiment. First, only items before the critical item were reliably depressed, and, second, as rate of presentation decreased, amnesia disappeared.

---

[1] A portion of this report is based on a dissertation submitted to the Department of Psychology, University of Alabama in partial fulfillment of the requirements for the PhD degree. Portions of the research were supported by a grant from the University of Dayton to the first author. I wish to thank Norman Ellis who suggested the second experiment, the reviewers whose very thorough efforts led to many useful suggestions for improving the original manuscript, and Kay Smith and her staff for their assistance.

[2] Requests for reprints should be sent to Douglas K. Detterman, Department of Psychology, University of Dayton, Dayton, Ohio 45409.

Rehearsal.—Jenkins and Postman (1948) felt that the outstanding qualities of the critical item withdrew attention from surrounding homogeneous items which led to reduced rehearsal and poorer retention of those items.

Total time.—A position similar to the rehearsal hypothesis was suggested by Waugh (1969) and states that total retention is a function of length of time allowed for rehearsal. Though rehearsal may be differentially distributed across items within a list, total retention is constant across lists. Unlike the rehearsal hypothesis, the total time hypothesis implies that superior retention of the critical item will be compensated for by a deficit in retention of surrounding items resulting in equivalent retention of lists with and without critical items. The rehearsal hypothesis does not necessarily imply compensation to the point of equivalent retention.

Rehearsal and intake.—Ellis et al. (1971) rejected the consolidation hypothesis because of two findings inconsistent with the position. Their experiments demonstrated anterograde amnesia and suggested item rather than time dependence of the effect. To account for anterograde amnesia by consolidation processes would require that the critical item act forward in time. Both retrograde and anterograde amnesia were thought to be item dependent because the same number of items were included in the effects regardless of rate of presentation. In opposition to this finding, consolidation processes are usually viewed as time dependent. Alternate and separate explanations were advanced for anterograde and retrograde amnesia by these investigators. The rehearsal hypothesis was employed to account for retrograde effects. This explanation did not appear appropriate for the anterograde effect since use of a recognition task where rehearsal was thought to be minimal eliminated retrograde but not anterograde amnesia. Increase in magnitude of amnesia with faster rates of presentation suggested an encoding deficit. It was postulated that recirculation of the critical item through the limited capacity primary memory preempted intake of the following items.

Retrieval and transfer.—On the basis of a review of the literature concerning clinical amnesia, Ervin and Anders (1970) conclude that retrograde clinical amnesia results from retrieval failure because affected patients demonstrate "islands" of intact memory, recall missing information under certain conditions, and often regain "lost" memories with passing time. In cases of anterograde amnesia, sensory and primary memory processes appear normal but attempts to recall material which has been "out of mind" for any appreciable period generally fail. This observation suggests a deficit in the ability to transfer material from primary to secondary memory.

In summary, induced retrograde amnesia can be viewed in terms of consolidation, rehearsal, or retrieval processes. Anterograde amnesia could involve rehearsal, encoding, or transfer processes. The present experiments tested these possiblities.

EXPERIMENT I

If induced retrograde or anterograde amnesia is related to rehearsal, those operations which attenuate rehearsal should also reduce the amnesic effect. In this experiment faster rates of stimulus presentation and instructions not to rehearse were used to minimize rehearsal.

Method

Subjects and procedure.—College students from eight psychology classes served as *S*s. The four treatment groups shown the first set of lists were composed of students enrolled in introductory psychology courses. A replication, with a second set of lists, used students in upper-level psychology courses. Each group participated during its regular class period and all students who reported for class were tested. To facilitate analyses, *S*s protocols were randomly selected in order to constitute groups of 40 each. All *S*s were naive except 9 who had participated in a similar study.

Twenty different lists of 15 items each were presented in a free-recall paradigm. All items in the list, except the critical item, were line drawings of common objects (e.g., hat, pill, pencil sharpener). Of the 20 lists, 10 were experimental and contained a critical item at the eighth serial position. The critical items were photographs of groups of male and female nudes, with exposed genitalia, taken from "sunbathing magazines." The experimental lists containing critical items were selected randomly with the following restrictions: a critical item could not appear on the first trial; no more than three successive trials could be composed entirely of homogeneous items; and no more than three successive trials could contain a critical item. A critical item appeared in Lists 3, 5, 6, 9, 11, 14, 15, 18, 19, and 20. The remaining 10 control lists were composed entirely of line drawings of common objects.

The stimulus lists were photographed on 16-mm. film, one stimulus item per frame, with additional frames inserted as a ready signal before the onset of each list and after each list as a signal for recall. During testing, the film was shown to each class with a projector (L-W Model 800), modified to allow the presentation of a single frame for an interval specified by *E*. The presentation of each of the 20 lists, then, was as follows: three frames for the word "Ready," the presentation of 15 stimulus items, and five frames of the word "Recall." On the third "Recall" frame, the projector was stopped for 1 min. to allow for recall. The *S*s wrote their responses on prepared answer sheets in any order they wished.

The design employed four experimental groups in a 2 × 2 × 2 × 15 complete factorial. Rate of presentation (.75 or 1.50 sec. per item), rehearsal instructions (instructed or not instructed), type of list (control lists composed entirely of homogeneous items or experimental lists containing a critical item at Serial Position 8), and Serial Position (Serial Position 1–Serial Position 15) were the variables of primary concern. Rate of presentation and rehearsal instructions were varied between *S*s. The others were varied within *S*. All *S*s were given standard free-recall instructions. In addition, that half of the *S*s given rehearsal instructions were told to think only of the item being presented and no other.

Results

Scores for analyses were number of correct responses at each serial position. Though no instructions were given regarding the correct response to be used for critical items, *S*s seemed to have no trouble in selecting an appropriate label. Responses given to indicate the critical item ranged from "people" to more descriptive phrases. Since a preliminary analysis showed that the amnesic effect decreased over lists, only the first five lists for each type of material were considered in subsequent analyses. The attenuation over lists was probably due to adaptation to the critical item and influenced all conditions equally so group relationships were unchanged. Analyses were conducted separately on items occurring before (Serial Positions 1–7) and after (Serial Positions 9–15) the critical item, excluding the critical item. Serial Position 8 was excluded from analyses to preclude interactions irrelevant to the assessment of the amnesic effect which would be created by superior retention of the critical item. Analyses were conducted separately on positions before and after the critical item because anterograde and retrograde amnesia may be related to different factors. In the following analyses, the replication involving different lists was included as a factor but was not involved in any significant effects.

In some of the previous work concerning induced amnesia, researchers have statistically demonstrated the effect by testing for significant differences between the single item preceding or following the critical item and the corresponding item in the control list. This procedure amounts to analyzing an interaction before it has been shown to exist. Although this approach seems justifiable on the basis of a priori hypotheses concerning induced amnesia, a statistically more conservative tack would be to demonstrate that items before or after the critical item are less well remembered, on the whole, than corresponding items in control lists. In addition, the closer an item is to the critical item, the larger the amnesic effect should be. The Serial Position × Type of List interaction best represents this state of affairs. Though only the Serial Position × Type of List interactions are reported, in every case the results are consistent with less conservative comparisons. For anterograde and retrograde effects and their interaction with other variables, conclusions based on an analysis of the item or two items most adjacent to the critical item agree perfectly with those based only on the appropriate interaction.

Figure 1 shows the effects of the two variables of principal concern, rehearsal instructions and rate of presentation. Both variables seemed to have the expected effects of limiting rehearsal as evidenced by inferior recall over the first seven serial positions with the faster rates of presentation, F (1, 312) = 31.5, $p < .01$, and when *S*s were instructed not to rehearse, F (1, 312) = 5.5, $p < .01$. As demonstrated by a Serial Position × Type of List interac-

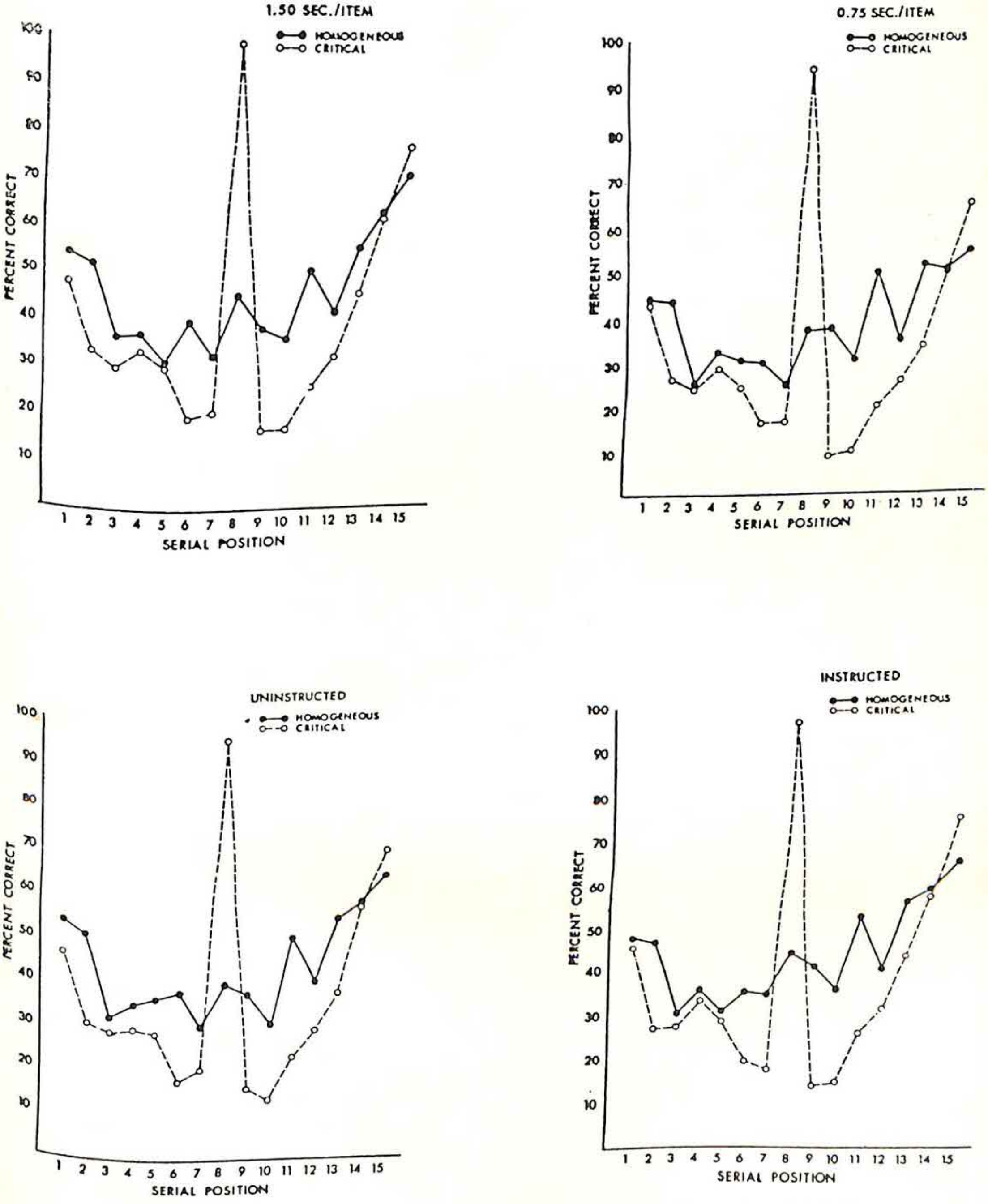


FIG. 1. Percent correct recall by serial position as a function of rate of presentation and rehearsal instruction each collapsed over the other.

tion, both retrograde, F (6, 1,872) = 22.6, $p < .01$, and anterograde, F (6, 1,872) = 69.1, $p < .01$, amnesia are present and substantial.

Instructions to limit rehearsal had no effect on retrograde amnesia. They did, however, influence anterograde amnesia as indicated by the Serial Position × Type of List × Rehearsal Instructions interaction, F (6, 1,872) = 2.3, $p < .05$, but in the direction opposite to that which would be predicted from a rehearsal hypothesis. The *S*s instructed not to rehearse demonstrated more severe anterograde amnesia. This effect was greater for serial positions adjacent to the critical item.

The Serial Position × Type of List × Rate interaction, F (6, 1,872) = 3.0, $p < .01$, indicated that slower rates of stimulus presentation led to more profound retrograde amnesia. This finding is inconsistent with Tulving's (1969) results and is opposite to predictions made on the basis of a consolidation hypothesis. The effect of rate of presentation on anterograde amnesia is only marginal, F (6, 1,872) = 2.0, $p < .06$, but it would seem that it is just the opposite of that found for retrograde amnesia. Faster rate induces more severe anterograde amnesia.

According to the total time hypothesis, lists containing a critical item should be recalled, overall, as well as those not containing a critical item. This was not the case. Lists of homogeneous items showed an average total recall of 43% while for lists with critical items recall was only 37%, a highly significant difference, F (1, 312) = 284.2, $p < .001$.

A separate analysis was conducted on Serial Position 8 and showed that retention of the critical item was in part, dependent on rate of presentation. The slower rate of presentation led to superior retention of the critical item.

EXPERIMENT II

Experiment I suggests that retrograde amnesia becomes more severe with slower rates of presentation while the opposite is true for anterograde amnesia. Faster rates lead to more severe anterograde effects. All of the experiments in which induced amnesia has been demonstrated have confounded the presentation rate of list items with the exposure duration of the critical item. It seems possible that the magnitude of the two amnesic effects are really expressions of these separate variables. In this experiment exposure duration of the critical item was manipulated independently of presentation rate of the list items.

Method

Subjects and procedure.—The *S*s were 80 volunteers from psychology classes who participated in groups of 6 or fewer. Half of the *S*s were used to replicate the experiment with a second set of lists.

The apparatus and procedure were the same as in Exp. I with two exceptions. First, only 11 lists were presented. The first list was for practice and was not scored. Half of the remaining 10 lists contained a critical item at the eighth serial position, as in Exp. I. Second, a modification of the apparatus was made so that exposure duration of the eighth serial position could be fixed independently of presentation rate of other list items.

The design of the experiment was a 2 × 2 × 2 × 15 complete factorial. Rate of presentation of list items appearing in Serial Positions 1–7 and 9–15, (.5 or 3.0 sec. per item) and exposure duration of items at the eighth serial position (.5 or 3.0 sec.) were between-*S* variables. Serial position and type of list (control lists composed entirely of homogeneous items or experimental lists with a critical item in Serial Position 8) were within-*S* variables. In this experiment, no instructions were given regarding rehearsal.

Two of the four groups formed by this design saw all items at the same rate of either .5 or 3.0 sec. per item. The other two groups saw the item at the eighth serial position for a duration different from other list items. Note that for these groups any item, either a critical or a homogeneous item, was shown for a different duration if it appeared in the eighth serial position. This was done to provide control for differences in performance that might result from changes in rate. One of these groups, then, saw the eighth item for 3.0 sec. and the list items at a rate of .5 sec. per item. For the other group the rates were reversed.

Results

Analyses were conducted, as in Exp. I, on the number of items correctly recalled at each serial position and separately for those items before and after the critical item. As in Exp. I, no significant differences resulted from the use of two different lists.

Figure 2 shows the results for each group. Again, both retrograde and anterograde amnesic effects are present and substantial. Side-by-side comparisons show the effects of lengthening exposure duration of the critical item while presentation rate of list

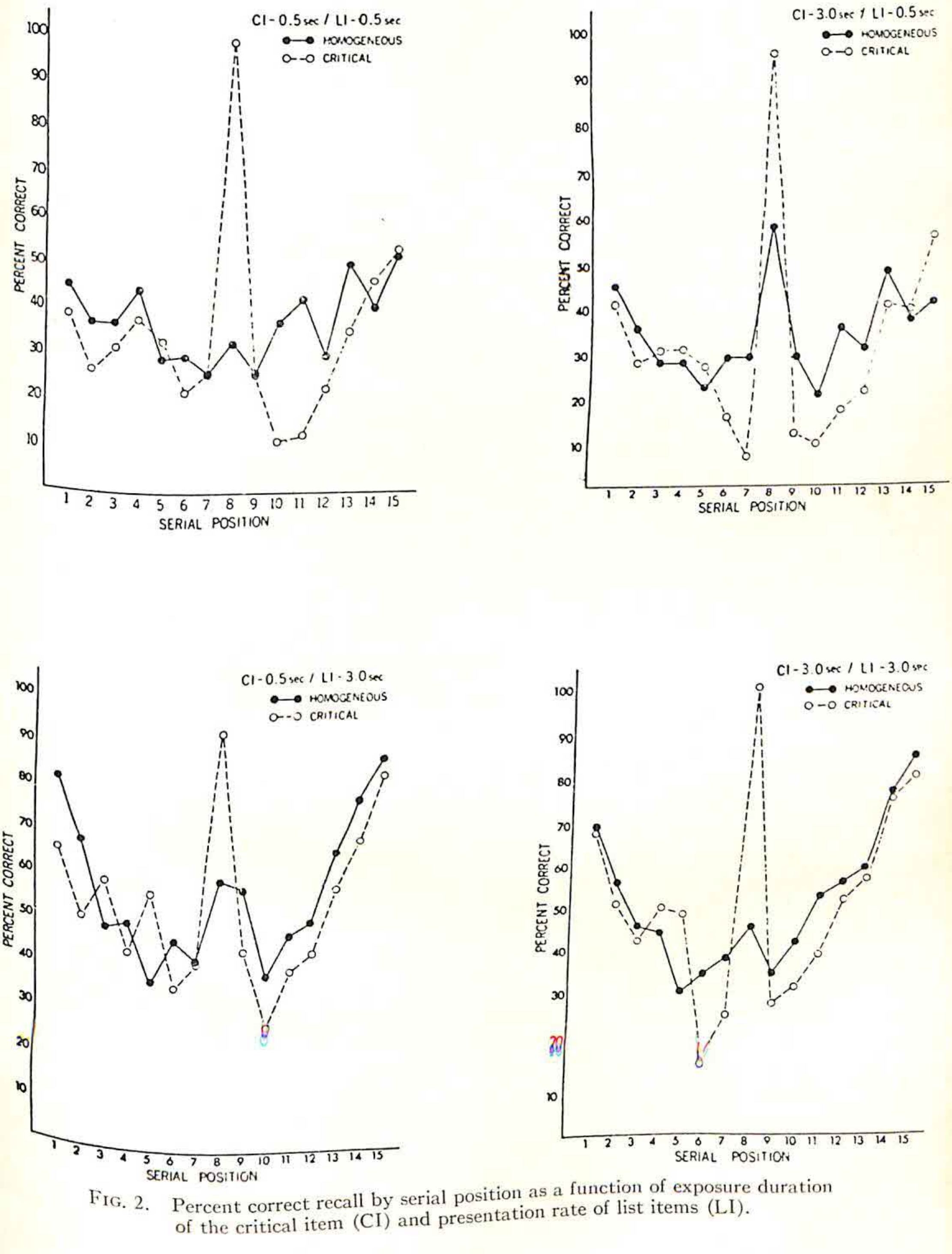


FIG. 2. Percent correct recall by serial position as a function of exposure duration of the critical item (CI) and presentation rate of list items (LI).

items is held constant. Top to bottom comparison shows the results of lengthening rate of presentation of list items while holding exposure duration of the critical item constant. These comparisons clearly demonstrate that retrograde amnesia grows more severe with increases in exposure duration of the critical item but that anterograde amnesia is lessened by increasing rate of presentation of list items. That exposure duration of the critical item is the sole determinant of retrograde amnesia is indicated by the Exposure Duration × Type of List × Serial Position interaction for the first seven serial positions, F (6, 432) $= 2.4$, $p < .05$. Presentation rate of list items is not involved significantly in any of the interactions which define retrograde amnesia. Similarly for anterograde amnesia only presentation rate of list items influence the magnitude of the effect as indicated by the Presentation Rate × Type of List × Serial Position interaction, F (6, 432) $= 2.3$, $p < .05$. Again no interaction defining the effect and involving exposure duration of the critical item for the last seven serial positions is significant. Unlike the results of Exp. I, retention of the critical item is not directly related to magnitude of the retrograde effect.

DISCUSSION

Two findings from these experiments bear most directly on explanations of both retrograde and anterograde induced amnesia. First, limiting rehearsal had no predictable influence on the amnesic effect. In Exp. I, instructions to limit rehearsal appeared to be effective because *S*s who were so instructed remembered fewer early items than those who were not. This result would be predicted from most theories suggesting rehearsal as an important process in memory. Though rehearsal seemed effectively limited, this had no major influence on induced amnesia. Although when rehearsal was minimized, anterograde amnesia did become more severe, this result would not be predicted from a rehearsal hypothesis. It might be that *S*s attempt to follow instructions to minimize rehearsal further complicated an already difficult task. This additional complication may well account for the larger amounts of induced anterograde amnesia demonstrated by *S*s instructed not to rehearse. Certainly if rehearsal processes were involved in the amnesic effect, changes in those processes should have resulted in more substantial differences than those found here. For this reason, it seems unlikely that either retrograde or anterograde induced amnesia can be accounted for by any hypothesis based on rehearsal processes.

This argument also eliminates the total time hypothesis as an explanation of either induced retrograde or anterograde amnesia, since the empirical relationship that equivalent study times result in equivalent retention regardless of the distribution of that time is generally accounted for on the basis of rehearsal processes. The present findings also offer an exception to the empirical relationship. When lists were presented for the same amount of time, lists not containing a critical item were better remembered than those containing one.

The second finding having implications for theoretical explanations of the amnesic effect is that anterograde and retrograde amnesia are determined by different variables. This would seem to suggest that anterograde and retrograde amnesia may be related to different processes.

Induced Retrograde Amnesia

Severity of retrograde amnesia is not rate dependent but directly related to exposure duration of the critical item. In Exp. I, amount of retrograde amnesia was dependent upon rate of presentation. However, in Exp. II, when exposure duration of the critical item was manipulated independently of the presentation rate of list items, it was found that retrograde amnesia depended only on exposure duration of the critical item. Rate dependence is central to the consolidation hypothesis suggested by Tulving (1969), and since it does not appear to hold true, this explanation seems inadequate as an account of processes involved in retrograde amnesia.

The retrieval hypothesis is the only remaining explanation considered which is capable of accounting for retrograde amnesia. Though exclusion of other hypotheses may not provide strong support, there are three other pieces of evidence that make this position more compelling. Amnesic severity related to retention of the critical item, the constant number of items involved in the effects in a given experiment, and failure to demonstrate retrograde amnesia using a recognition task would all seem to argue for retrieval failure as the origin of retrograde amnesia. That amnesic severity is related to retention of the critical item is indicated by two findings from these experiments. In Exp. I, the better the critical item was retained the greater was the amnesic effect, while in Exp. II exposure duration of the critical item was related to the magnitude of retrograde amnesia. For reasons cited earlier, disruption of rehearsal or consolidation processes do not appear to be appropriate explanations for these effects. This would seem to rule out any reasonable explanations we can suggest which involve encoding or storage mechanisms, leaving only retrieval. Ellis et al. (1971) point out that retrograde amnesia has influenced a fixed number of items regardless of severity of the effect. This observation is supported by the present findings. In both experiments it would appear that only the two items before the critical item are involved in the amnesic effect. Since encoding processes seem to be more time-dependent than retrieval processes, again retrieval processes seem the best explanation. Ellis et al. (1971) also found no retrograde amnesia when a recognition task was employed. Though they accounted for this finding in terms of minimal rehearsal demands in the recognition task, as good a case could be made for minimal retrieval demands. Though any one of these three findings may not, by itself, make a strong case for retrieval failure as the origin of retrograde amnesia, taken together they are difficult to explain otherwise.

Though these findings may implicate retrieval processes as important in the origin of induced retrograde amnesia, they do not provide a mechanism for the effect. It may be that the distinctive qualities of the critical item interfere with retrieval of preceding items in some way. This position seems very similar to the hypothesis recently advanced by Meyer (1972) to account for the effects of ECS on the recall of long-term habits. Rather than acting as massive punishment or disrupting consolidation, Meyer views ECS as inhibiting access to still existing engrams.

Induced Anterograde Amnesia

With regard to induced anterograde amnesia, those hypotheses implicating encoding, which involves those processes which occur before storage, seem most reasonable since magnitude of the effect is directly dependent on rate of presentation of list items. Ellis et al. (1971) suggest that the critical item preempts primary memory and limits the intake of successive items. Ervin and Anders (1970) believe that clinical anterograde amnesia reflects a failure to transfer items from primary to secondary memory. If the Ervin and Anders hypothesis is relevant, it should not be possible to demonstrate induced anterograde amnesia in primary memory as Ellis et al. would predict. The only evidence bearing on this point comes from experiments by Tulving (1969) and Saufley and Winograd (1971). They introduced critical items in Serial Position 14 of a 15-item list. These last items in a list are generally thought to be fairly pure reflection of primary memory. In three of the five cases reported, memory of the fifteenth item was better than in lists with no critical item. For the other two cases it appeared to be close to the same. These results suggest that induced anterograde amnesia may not be demonstrable in primary memory and seem to favor the hypothesis that anterograde amnesia results from a failure to transfer items from primary to secondary memory as Ervin and Anders suggest. It should be pointed out that in neither of these studies were consistently reliable anterograde effects found. In any event more explicit tests of the two positions are required.

These experiments suggest that the causes of induced retrograde and anterograde amnesia are different. Retrograde amnesia appears to result from an inability to retrieve stored information and anterograde amnesia from an encoding failure. A major question left unanswered is why these results differ from those obtained by others (e.g., Tulving, 1969). The most obvious difference between experiments has been the type of critical item used. Photographs of nudes are different in many ways from the verbal materials which have more commonly been used to demonstrate induced amnesia. Perhaps an explanation of what makes something outstanding would also lead to an understanding of the processes involved in induced amnesia.

REFERENCES

Ellis, N. R., Detterman, D. K., Runcie, D., McCarver, R. B., & Craig, E. M. Amnesic effects in short-term memory. *Journal of Experimental Psychology*, 1971, **89**, 357–361.

Ervin, F. R., & Anders, T. R. Normal and pathological memory: Data and conceptual scheme. In *The neurosciences second study program*. New York: Rockefeller University Press, 1970.

Jenkins, W. O., & Postman, L. Isolation and "spread of effect" in serial learning. *American Journal of Psychology*, 1948, **61**, 214–221.

Meyer, D. R. Access to engrams. *American Psychologist*, 1972, **27**, 124–133.

Saufley, W. H., Jr., & Winograd, E. Retrograde amnesia and priority instructions in free recall. *Journal of Experimental Psychology*, 1970, **85**, 150–152.

Tatuno, T. On the effect of the isolated term and the identical term in serial learning. *Japanese Psychological Research*, 1961, **3**, 17–27.

Tulving, E. Retrograde amnesia in free recall. *Science*, 1969, **164**, 88–90.

Wallace, W. P. Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. *Psychological Bulletin*, 1965, **63**, 410–424.

Waugh, N. C. Free recall of conspicuous items. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 448–456.

(Received December 27, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 317–326

ENCODING AND RETRIEVAL IN VISUAL MEMORY TASKS[1]

NANCY FROST[2]

University of Washington

Four studies examined memory for visually and semantically categorized pictures. Stimuli were designed so that recall organization could be based on one, the other, or both types of categories. Free recall of picture names and recognition were performed by *S*s who expected either recall or recognition memory tests. Those *S*s expecting recognition recalled by *combining* visual and semantic categories and could efficiently—in terms of RT—perform a visual recognition task. The *S*s who expected recall, however, clustered by semantic categories only and were efficient only in name recognition. It was concluded that pictures are encoded differently depending on task expectation. Parallel access of visual and semantic memory codes occurs; but when recognition is expected, a visual cue provides faster access, and when expecting recall, verbal access is more efficient.

Typically, psychologists have studied the organization of long-term memory (LTM) in terms of verbal stimuli. Bousfield (1953) noted that, left to his own preferences, *S* would recall a list of words in groups (or *clusters*) of semantically similar items. Others (see Shuell, 1969) have noted both category and word-association organizations in *S*s' free recall, while recall is seldom arranged by parts of speech (Cofer & Bruce, 1965) or by geometric similarities designed into nonsense forms (Bousfield, Berkowitz, & Whitmarsh, 1959). The accumulation of evidence suggests that LTM is semantically organized (Bower, 1970; Mandler, 1967). Adult *S*s retrieve words according to their associative, categorical, and even idiosyncratic meanings (Bousfield & Bousfield, 1966; Cohen, 1963; Mandler, 1967; Tulving, 1962); and *S*s can remember nonverbal stimuli, e.g., nonsense shapes, better when those stimuli have descriptive verbal labels to aid encoding or retrieval (Ellis, 1968).

Several studies have attempted to investigate organization effects in the memory for visual stimuli. Bousfield et al. (1959) used a clustering paradigm to study free reproduction of complex geometric shapes. The shapes were categorized by arbitrarily selected geometric subshapes. Their *S*s did not cluster according to the shape categories until several study-test trials had been performed. The authors' conclusions imply a verbal recoding hypothesis of the visual elements in each stimulus. Byrne and Brooks (1970) looked at ordered, rather than free, recall of the names of objects in picture composites. The objects were organized into a scene where the spatial organization of items conflicted with a categorical arrangement of item names. They found evidence that *S*s searched an image of the composite while classifying names. Two studies have used pictorial stimuli in clustering paradigms. Scott (1967) demonstrated that pictures can indeed be recoded and organized according to their verbal labels. Frost (1971) used pictures of common objects categorized not by semantic characteristics, but by the visual similarities of the pictures themselves. Her *S*s clustered according to the similar shapes of the drawings while free recalling the names of the pictured objects.

The latter study (Frost, 1971) emphasizes a notion overlooked by some researchers—if we are capable of remembering large amounts of visual information (as in the picture recognition studies of Shepard, 1967; Standing, Conezio, & Haber, 1970), there may be an appropriate visual means of organizing and thereby retrieving that information from memory. Certainly the Byrne and Brooks (1970) and the Frost studies indicate that given no choice in the matter, *S*s can and do organize their recall from LTM according to spatial relationships within or among the stimuli. A verbal organization hypothesis seems insufficient to explain these results.

---

[1] Experiment I, II, and III are based on research conducted at the University of Oregon. Funds for *S*s and computer use were provided by the Advanced Research Projects of the Department of Defense, monitored by the Air Force Office of Scientific Research under Contract F44620-67-0099, principle investigator, Ray Hyman. The support for Exp. IV was provided by the United States Air Force Office of Scientific Research, Air Systems Command Grant AFOSR 70-1944 to Earl B. Hunt, University of Washington.

[2] Requests for reprints should be sent to Nancy Frost, who is now at Department of Psychology, Princeton University, Princeton, New Jersey 08540.

In one sense, Frost's (1971) study avoids the question of semantic contributions to free-recall organization by giving *S*s no semantic list structure with which to deal; Byrne and Brooks (1970), too, forced their *S*s to recall according to spatial, not semantic, order. In another sense, however, the very fact that their stimuli were verbally encodable may explain the difference between the Bousfield et al. (1959) findings and Frost's visual clustering study.

The following experiments begin an investigation of the relationship between semantic and visual memory. Specifically, they attempt to study the contributions of semantic and visual codes to the memory for pictures which are categorized both by shape and by name. Questions of interest are three: (*a*) When memory is searched for doubly encodable items, does the search combine visual and semantic content or does organization along one stimulus dimension (either visual or semantic) preclude the use of other categorizations? (*b*) Are both visual and semantic codes for a single item stored in LTM? (*c*) If both codes are maintained in memory, how are they accessed? The questions are examined in terms of two retrieval processes, free recall and recognition, which typify the larger part of our daily memory tasks.

EXPERIMENT I

Where two schemes of organization exist in the stimulus set, there are three possibilities for those schemes to affect recall. Each possibility is the basis for a model of memory access in free recall to be explored in Exp. I.

First, one type of organization may be used exclusively of the other. When the visual organization scheme is ignored, then Model 1 represents the verbal organization hypothesis of visual memory. Second, memory may be searched first according to one organizing scheme, and later switched to the other scheme to pull even more items from memory. The alternating of search modes, Model 2, describes a retrieval system without communication between semantic and visual modes, with the capability for only unidimensional memory search. In Model 3, visual and semantic searches may be combined to retrieve items from memory. The latter model can be likened to the parallel perceptual scanning that Cohen (1970) has found for conjunctive (semantic and visual) targets, and to the retrieval process proposed in Wickelgren's (1970) multitrace theory of memory.

A free-recall paradigm was used, measuring the visual, semantic, and combined clustering for each *S*. Computer simulations of the three models estimated expected amounts of double (semantic and visual) clustering which were then compared to the performance of human *S*s.

Method

Subjects.—Sixteen *S*s, 8 men and 8 women, were run individually in Exp. I. Each *S* participated in one 45-min. session, for which he was paid $1.50.

Stimuli.—Sixteen names of common objects—4 animals, 4 articles of clothing, 4 vehicles, and 4 items of furniture—were selected from the 10 most frequent items in the Battig and Montague (1970) category lists. Each of the 16 objects was drawn in four orientations: vertical, horizontal, slanted left, and slanted right. While visual category membership was defined by each drawing's axis of orientation, it should also be noted that visual similarities within each category were maximized by using as much line-by-line redundancy in the drawings as possible (see examples in Fig. 1). The resulting 64 drawings were penned in black ink on 5 × 8 white cards, 1 per card.

A different combination of 16 drawings was then selected for each *S*. Each stimulus set included 1 drawing of each of the 16 objects, chosen so that the 16 drawings represented all four orientations equally. Further, within each four-item semantic category, two examples shared the same orientation; while the remaining two items shared a different orientation. Likewise, each four-item visual category represented two of the semantic groups with two items in each. An example of the overlapping of semantic and visual categories appears in Fig. 1.

Four semantic orders were selected by randomly pairing items from the semantic categories without replacement. Each semantic pair in these four orders was then represented in all four orientations to complete the selection of 16 different stimulus lists. An involved list selection process was used to eliminate biases due to unusually strong associations between any two items. Because the semantic and orientation pairings were selected without replacement, each semantic pairing shares the same orientation in only $\frac{1}{3}$ of the lists; and the same semantic pair differs in orientation $\frac{2}{3}$ of the time. Therefore, apparent visual clustering in recall protocols cannot be the total result of strong semantic associations and confounded semantic-orientation pairing.

Procedure.—The *S* began the experimental session by participating in a "practice" experiment as described in Frost (1971). Eight drawings, corresponding to none of the visual and semantic categories of the stimulus lists, were shown to *S* sequentially at a 1.5-sec. rate. The *S* was then shown a series of cards containing 3 drawings per card and asked to point to drawings which were identical to those from the memory set of 8. In this way, all *S*s were set to expect a recognition task involving the 16 stimulus drawings; and instructions read to *S* described the practice task as being a sample of the experimental task to follow.

The stimulus drawings were presented to *S*, one at a time, at a 1.5-sec. rate. Stimuli were randomized with the restriction that there could be no sequential repetitions of either shape or semantic categories. Next, *S* was handed a booklet which included instructions and materials to do a number-cancelling task. Work on this task filled a 15-min. retention interval between stimulus presentation and the free recall of picture names; free recall was then given orally by *S*, within a 3-min. time limit.

Results

Free-recall data for picture names were collected from each of the 16 *S*s. Free-recall protocols were scored for sequences of repeated visual categories, of repeated semantic categories, and of doubly repeated categories (i.e., where an item belongs to both the same visual and semantic categories of its immediate predecessor). The occurrence of two consecutive items with at least one (semantic or visual) category in common was scored as one *repetition* (*R*), and the total number of visual, semantic, double, and single category repetitions were recorded for each *S*. Intrusions of new items into the recall protocols were excluded from the repetition scoring and from the total number of words recalled. Details of the repetition scoring procedure can be found in Frost (1971).

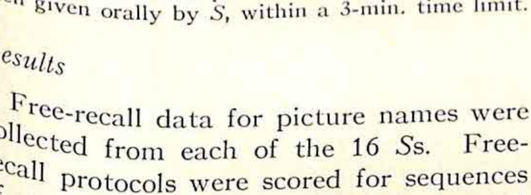


FIG. 1. An example of stimuli from Exp. I. (Semantic and shape categories are partially overlapped. PANTS-SWEATER represents a double category repetition; PANTS-SHOE and PANTS-ELEPHANT are single category repetitions.)

Clustering (*C*) measures were calculated for each *S* according to the *C* measure recommended by Dalrymple-Alford (1970):

$$C = \frac{\text{Observed } (R) - \text{Minimum } (R)}{\text{Maximum } (R) - \text{Minimum } (R)},$$

where minimum possible *R* is calculated $2 \cdot m_a - N - 1$, and maximum possible R is defined $N - n$, N representing the total number of words recalled, n the total number of categories recalled, and m_a the largest number of words recalled in a single category. The measure yields a clustering score that is independent of the number of words recalled and which can vary, for all *S*s, between .0 and 1. Mean *C* of .424 (visual) and .598 (semantic) obtained. Expected clustering scores, E(*C*), were also calculated for each *S* by substituting the expected[3] number of *R*,

$$\frac{\Sigma_i(m_i - 1)^2}{N - n},$$

into the equation for *C*. *T* tests performed on the difference between observed and expected scores yielded 95% confidence intervals of .173 ± .099 (for visual categories) and .370 ± .122 (for semantic categories). Because zero, the expected difference score for chance clustering, falls outside both confidence intervals, the null hypothesis (that clustering did not occur) is twice rejected. These results suggest that *S*s used both visual and semantic categories to organize their recall. Mean overall C, indicating the amount of combined visual and semantic organization in the free-recall protocols, was .51 which is well above the (theoretical) chance score of .33.

---

[3] The formula used is the estimated maximum variance of observed repetitions, given that the actual clustering level is zero. Under such conditions, it is expected that the observed repetitions are due wholly to chance, and the expected number of repetitions therefore approaches the theoretical chance level. A similar version of the formula has also been suggested by Roenker, Thompson, and Brown (1971).

The proportions of total *R*s which were double ($\bar{X} = .505$, variance $= .038$) were transformed by the Freeman-Tukey arc sine transform for samples of small size (Mosteller & Youtz, 1954), and the transformed scores were used to establish confidence limits for the true mean proportion of double R to total $R(D/T)$. Mean arc sine 45.65 and variance 83.72 estimate the true D/T to be between arc sine 38.95 and 52.35 with 95% confidence.

Simulation.—The three models outlined above predict different D/Ts, depending on the method of accessing the visual and semantic information for recall. The probability that an R is double, given that only one type of code guides recall, is $\frac{1}{3}$—in Fig. 1's example, PANTS can be followed by SWEATER, SHOE, or HAT if a semantic R occurs (or SWEATER, BEAR or ELEPHANT if orientation repeats). On the other hand, if Model 3 is correct, p(SWEATER) is 1.0; the next-recalled item, however, has a p(double R) = .0, since in free recall S typically "samples without replacement" from the stimulus set. For all three models, the probability that the next R repeats both semantic and visual categories changes with the number of items that have been recalled, the order in which items are recalled, and the degree to which S clusters. Defining equations which predict the relative D/Ts for each model becomes impossible; therefore, computer simulations of free-recall sampling were performed in order to establish the mean predicted D/T for each model as well as the variance associated with these predictions. A FORTRAN program was written for the DEC PDP-15 computer which employed a pseudo-random number generator and a recall selection rule to determine the order of simulated free recall. The selection rules were as follows:

1. *Single dimension:* Using only one type of categorization (e.g., semantic), select the item with the greatest probability of following the last recalled item.

2. *Switching:* Select the item with the highest probability of following the last recalled item, based on either semantic or visual category membership.

3. *Combinatory:* For each unrecalled item, multiply the p(recall) associated with its visual and semantic category membership. Select the item with the greatest combined probability.

One assumption is necessary to implement the simulation—that the probability of an item being followed by a member of the same (either semantic or visual) category is, on the average, higher than the probability that any nonrepetition will follow. This assumption is justified by the fact that semantic and visual category clustering do occur in the free-recall protocols. Given this one assumption, it is unnecessary to attach any numerical values to the probabilities of within- and between-category repetitions; and it is also unnecessary to stipulate whether the probability of a semantic versus orientation repetition is greater.[4]

Free recall for 1,000 *S*s was simulated for each model. Model 3 predicts a mean D/T of .533 with variance = .0. Model 1 and 2 predict proportions of .326 (variance = .017) and .282 (variance = .012), respectively. It should be noted that the simulations were designed to do something that human *S*s cannot: to correctly recall

---

[4] Let A = the probability (p) that a semantic repetition will occur; $B = p$ of a visual repetition; $a = p$ of a semantic nonrepetition; and $b = p$ of a visual nonrepetition. Given the assumption that $A > a$ and $B > b$, A and B can be rewritten as $A = a + k$ and $B = b + c$, where k and c are unspecified positive constants. Therefore, $A \cdot B = (a + k)(b + c) = ab + ac + bk + kc$, p of a double repetition. The p of a single repetition can be written $A \cdot b = b(a + k) = ab + bk$, or $B \cdot a = a(b + c) = ab + ac$. The p of a nonrepetition, $a \cdot b$, is simply ab. The expanded equations illustrate a greater probability of a double repetition than any single repetition or nonrepetition, under the combining rule of Model 3. This relationship holds regardless of the (positive) values placed on any variables; and the relationship obtains for association based models, where strengths of association among stimuli are added, rather than multiplying the p of possible occurrences.

all stimulus words and to cluster perfectly at all times. To ensure that the D/T simulations can be realistically compared to that of the experimental *S*s, two parameters were introduced into the program. First, the number of words recalled by each simulated *S* was limited to the mean number of words recalled by human *S*s (i.e., 10.0). Second, the level of clustering was reduced to that observed for experimental *S*s, 51%. The clustering level reduction was simulated by interrupting the selection process with an appropriate probability, .49.

With the addition of these two parameters, new D/T proportions were predicted for Models 1, 2, and 3 (1,000 *S*s per model). Model 1 predicts a D/T of .280, with variance .046; Model 2, a D/T of .267 with variance, .044; and Model 3, a D/T of .467 with variance .043. Respective arc sine transformations of these means are 32.00, 31.30, and 43.30. The confidence limits established for experimental *S*s $C(38.95 < \text{arc sine}\mu < 52.35) = 95\%$, clearly accept only Model 3 as an adequate fit to the data. It is felt that the two-parameter simulation gives the better D/T estimates because of the close correspondence to number of words recalled, number of repetitions made, and variance of the D/T scores; nevertheless, parameterless versions of the simulation programs also reject Models 1 and 2, while accepting Model 3. Model 1 is further rejected by the fact that both visual and semantic clustering were observed in experimental data.

Discussion

Experiment I dealt with grouped data, estimating a proportion, D/T, that is predicted for individual *S*s as well. While the grouped data fit Model 3, it may be the case that some of the *S*s behaved according to one model—and others, according to another or even none of the remaining models. To test for this possibility, a chi-square test for independence of D/T and individual *S*s was performed. Insignificance in this test, $\chi^2(15) = 5.36$, indicated that no difference is apparent in the performance between *S*s.

The results of Exp. I suggest, therefore, that *S*s were combining spatial and semantic information in order to recall picture names. The combining of codes is strong evidence that both kinds of codes are, in fact, stored for each stimulus drawing. Other experiments using recall and pictorial stimuli (see Bahrick & Boucher, 1968; Sampson, 1970) have produced results consistent with, but not requiring, a dual-encoding interpretation.

EXPERIMENT II

While it can be concluded from Exp. I that both visual and semantic codes are stored for each stimulus and that both codes contribute to free recall, the reason for the combining effect is unclear. It may be that the nature of the recall task dictates that *S* use all available information, and that other retrieval tasks, e.g., recognition, may not evidence a combination of visual and semantic information. Alternatively, the combining may be an encoding effect; accessing one type of memory code may necessitate the access of the other. Funkhouser (1968) and Wright and Kahneman (1971) suggest that task expectations affect the performance of memory tasks. Experiment II was designed to investigate the role of task expectation in determining the type of free-recall organization for doubly categorized stimuli. One group of *S*s replicated the procedure of Exp. I; while a second group was set to expect the name recall task which followed. If the combination of visual and semantic codes is characteristic of recall in general, then both groups should replicate the results of Exp. I. If, however, the combining effect is due to *S*'s expectation of the recognition task, then only the recognition-set *S*s should replicate the findings of Exp. I.

Method

Subjects.—Using the actual variance and simulated D/Ts from Exp. I, power was estimated for a reduced sample size. Power of .88 was considered sufficient to allow a sample of 12 *S*s per group in Exp. II. Therefore, 24 *S*s were selected by the University of Oregon Employment Service and were assigned by *E* to one of two groups. The groups were tested simultaneously, yoked pairs of *S*s receiving identical stimulus list presentations. Three men and nine women served in the recognition-set condition; four men and eight women served in the recall-set condition. Each *S* was run individually in one 45-min. session, for which he or she was paid $1.50.

Stimuli.—Stimuli and stimulus ordering rules were identical to those used in Exp. I. Three new pair-

TABLE 1

MEAN *C* CLUSTERING SCORES: EXPERIMENT II

Types of clustering	Recognition-set *Ss*		Recall-set *Ss*	
	Expected clustering	Observed clustering	Expected clustering	Observed clustering
Visual categories	.254	.537	.215	.284[b]
Semantic categories	.252	.684	.215	.593
Over all	.33[a]	.652	.33[a]	.522

[a] Theoretical chance level of clustering.
[b] *ns.*

ings without replacement of stimulus names were made, and each of these three lists of names was expanded into four, resulting in 12 different stimulus sets in which no orientation-name pair was repeated.

Conditions.—Two conditions were used in Exp. II. In one, *S*s were set to recognize pictures before proceeding with stimulus presentation and subsequent picture name recall. In the other, *S*s were set to expect the free-recall task which followed.

Procedure.—Two changes were made from the procedure reported in Exp. I. First, the recall-set *S*s were asked to free recall practice stimuli by name, instead of performing a forced-choice recognition task. Second, a forced four-choice recognition task followed free recall of the stimulus names. Each of the 16 stimulus objects was drawn in all four orientations on a single 5 × 8 card. The *S* was asked to point to the original stimulus drawing when shown each card in random order.

Results

The mean number of words recalled by recognition-set and recall-set *S*s were 10.7 and 10.3, respectively. The levels of visual, semantic, and overall clustering for both groups are presented in Table 1. The 95% confidence intervals for $C - E(C)$ scores indicate that recognition-set *S*s clustered by both orientation, $\mu = .286 \pm .154$, and name, $\mu = .429 \pm .163$; while recall-set *S*s clustered only by name, $\mu = .376 \pm .184$ (95% C for orientation includes zero). These results reject Model 1 for recognition-set *S*s and Models 2 and 3 for the recall-set group.

The D/T scores were transformed to arc sines by the Freeman-Tukey method, and confidence limits established for the population mean in each condition. Recognition-set *S*s, with mean D/T of .427 (arc sine data: 95%C = $37.599 < \mu < 45.329$) replicates Exp. I by rejecting Models 1 and 2, while accepting Model 3.[5] Recall-set data, however, differ from those of the recognition-set group. Mean D/T for the recall condition is .230, with confidence limits about the arc sine of 6, 95% C = $24.142 < \mu < 37.238$, which rejects Model 3.

Recognition data resulted in mean errors of 2.75 and 4.67 for recognition-set and recall-set conditions, respectively. Chance performance is 75% errors, or 12.

Discussion

The failure of the recall condition to match Model 3 indicates that the combined retrieval effect is not a requirement of the free-recall process. Instead, Exp. II suggests that the combination of visual and semantic content is in some way established at encoding time. It is not, however, a mandatory function of the encoding of pictorial stimuli, since different task expectations led recall-set *S*s to use only the semantic retrieval scheme. Recall-set *S*s do appear to maintain both visual and semantic codes in memory, as indicated by their orientation recognition performance and semantic clustering.

There are at least two ways in which input can influence subsequent retrieval from memory. First, the type of organization noted in free recall may be a reflection of differential storage structure. On the other hand, item storage could be identical for both groups, but the recall-set group may store a plan to recall according to names, perhaps rehearsing the stimuli differently than do recognition-set *S*s. Introspections of the recall-set *S*s deny the use of a better rehearsal strategy. They, just as most recognition-set *S*s, reported only that they covertly named each drawing once as it was presented. The *E* did note, however, that some recall-set *S*s used more distinctive labels for a few of the drawings than did recognition-set *S*s (e.g., TEDDY-BEAR instead of BEAR and LOCOMOTIVE instead of TRAIN).

EXPERIMENT III

Experiment III was designed to answer the question: Is the combination effect, found for recognition-set *S*s, due to a recall strategy adopted by *S*s "caught off guard" by the free-recall task, or is it a general characteristic of their ability to remember the picture stimuli? A reaction time (RT) task was used to discover whether the recognition-set *S*s also access visual information along with the name code in a (name) recognition task. Recall and recognition sets were used, as in Exp. II. If the performance of the two conditions differ within a recognition paradigm, then the combining effect must relate to a general retrieval process, encompassing both free recall and recognition.

---

[5] The variance for the recognition-set group in Exp. II, .016, is lower than both those calculated for Exp. I and predicted by the simulation of Model 3. It should be pointed out, however, that the Exp. II *S*s clustered more than Exp. I *S*s; therefore, a resimulation of Model 3 with an increased clustering parameter would predict a lower variance, as observed.

Experiment III substituted a name-recognition task for the free recall in Exp. II's paradigm. The *S* was required to decide whether the object pictured in a test stimulus had also been pictured in the set of remembered stimuli. RT to two types of YES responses could be measured (*a*) to stimuli identical to the original stimulus drawing, and (*b*) to drawings of an object which had been redrawn in a different orientation. In a name-recognition task, it would be to *S*'s advantage to ignore visual information (which is often misleading) and to access the name memory code directly. If RT to both types of YES response are the same, then *S* must be accessing directly the name code in both cases. If, however, the second type of YES response is slower than the first, then *S* may be accessing the name information in combination with visual codes. Evidence that orientation affects the recognition task would then suggest that the combining of visual and semantic codes during retrieval is not something that *S* can do, but something that he does because of the way the stimuli have been encoded.

Method

Subjects.—Sixteen *S*s were run individually in one hour-long session each. Three men and five women were assigned to each of the recognition and recall-set conditions. Yoked pairs of *S*s were run successively, each member of a pair receiving the same stimulus set as his partner. One additional *S* was run but was dismissed before completing the session; an approximate error rate of .5 (chance) left her RT data uninterpretable.

Stimuli.—Eight stimulus orders were selected at random from the 12 used in Exp. II. However, 7 substitutions were made in the total set of 64 pictures. Since the previous experiments attempted to measure visual clustering, visual similarities were maximized across semantic categories as well as within. On a few occasions, details of a single object were changed from one visual category representation to the next, in order to maximize similarities across objects (e.g., the CAR was drawn as a sedan in the vertical category but as a convertible in the slant categories). In Exp. III, however, the name-recognition task required all CAR drawings to be of the same convertible from different viewpoints, etc. The substitutions eliminated those error responses due to disagreement concerning verbal labels, rather than memory failure.

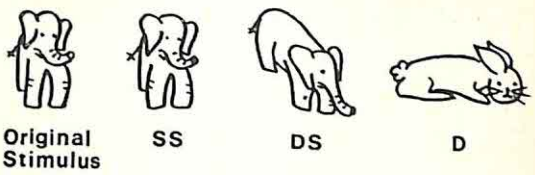


FIG. 2. An example of the test items used in Exp. III and Exp. IV recognition tasks.

Procedure.—The practice experiment, stimulus presentation, and first 10 min. of the interpolated task proceeded as described for Exp. II. The final 5 min. of the interpolated task was performed with pairs of numbers presented tachistoscopically. The *S* was asked to press a YES key with his left index finger if the pair were identical; he pressed a NO key with his right index finger if the numbers were different. The *S* received 36 such practice trials before beginning the RT recognition task which followed.

Before the recognition task, *S* was instructed to respond YES via a keypress when a stimulus occurred which had the same *name* as a drawing from the memory set of 16. The *S* saw each original stimulus drawing twice (32 Same-orientation, Same-name or SS stimuli); each stimulus object in two new visual categories (32 Different-orientation, Same-name or DS items); and 64 drawings of "new" objects—4 animals, 4 items of furniture, 4 items of clothing, and 4 vehicles, each drawn in the four orientations. Examples of the SS and DS drawings appear in Fig. 2. The 128 test cards were randomized for each yoked *S* pair and presented tachistoscopically until *S* responded. The *E* changed cards between trials, noted *S*'s response and RT, and gave *S* no feedback about his or her performance.

Results

The YES RTs were averaged by type for each *S*. Four mean YES times resulted: mean RT for the first and the second occurrence of each SS and each DS test card. Difference scores were then established for each *S* by subtracting the DS mean from the SS mean for both first and second occurrences or *trials*, and analysis of variance was performed on the differences, contrasting groups (retrieval-set condition),

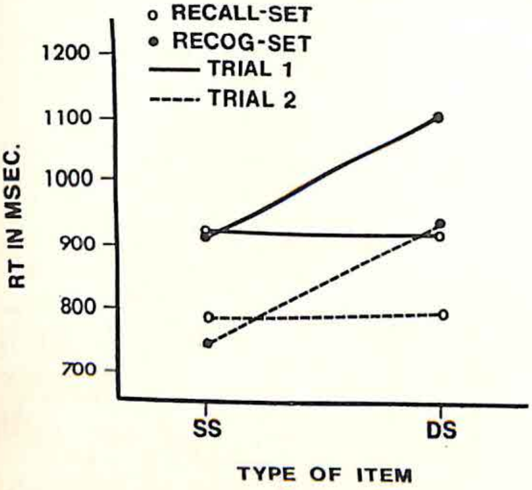


FIG. 3. Mean RT for YES responses in the name-recognition task, Exp. III.

trials, and *S*s. The source groups was significant beyond the .005 level, F (1, 14) = 14.139, indicating that the recognition-set *S*s took reliably longer to make DS judgments than did recall-set *S*s. The DS and SS RTs, graphed by trial and condition, are presented in Fig. 3. Neither trials nor interaction sources were significant. A practice effect is apparent in Fig. 3 that was not tested by analysis of variance.

Errors, for the most part, were not random but meaningful—e.g., when CAR was incorrectly called "new," it was usually called new more than once (maximum four errors per object). Errors occurred to 10.5% of the "old" test items for recognition-set *S*s, and to 11.5% of the old items for recall-set group. For recognition- and recall-set *S*s, respectively, 9.8% and 8.9% of the total responses were in error.

An analysis of variance was performed on *misses*, those SS and DS items incorrectly called new. In a Subject × Group × Type-of-Item analysis, only type-of-item attained significance, F (1, 14) = 15.93, $p < .005$. Both groups of *S*s made almost twice as many DS as SS errors ($\bar{x} = 2.75$ to 4.25, DS:SS, for the recognition-set *S*s and 2.25 to 5.00 for recall-set *S*s).

The RTs to new test items, since they contribute no information about the relationship between memories for form and picture names, are not considered in the present discussion.

Discussion

The RT differences noted for recognition-set *S*s indicate that the visual content of their memory was accessed in a task requiring only name matches to the remembered stimuli. As in the previous experiments, these *S*s seem to be combining visual and semantic codes to perform the memory task; however, in Exp. III the combining takes on an obligatory characteristic, since the longer RT to DS items suggests that orientation interfered with the speeded judgment task. Recognition-set *S*s typically reported that they "responded YES if the picture *looked* enough like something [they] remembered—if it looked very different, [they] couldn't decide."

For recall-set *S*s, on the other hand, name matching was equally efficient for both kinds of test items, suggesting that name codes were accessed to make the match. Although orientation did not interfere with the speed of judgment, fewer errors for DS items suggests that the visual information did contribute to the accuracy of the judgment.

EXPERIMENT IV

If both name and spatial code are accessed in both retrieval-set conditions in Exp. III, then the data suggest that the two codes are accessed in parallel—with more efficient access for recall-set *S*s via name code, and for recognition-set *S*s via visual code. The difference in number of DS and SS errors, suggestive of dual access for recall-set *S*s, does not require the dual-access explanation. It may be that these *S*s create such a distinctive verbal name code for each stimulus that DS items are named differently than are their SS counterparts; only an SS test item, then, would be recognized accurately. Experiment IV was designed to provide a stronger test of the hypothesis that recall-set *S*s actually store and can access the visual information contained in the pictured stimuli. Experiment IV's design called for a replication of Exp. III with the addition of visual memory judgments to the name-recognition task; *S*s who were set to recall and those set to recognize responded YES only to those test items which were identical to the remembered stimuli (SS items). If DS items produce slower NO RTs than do "new" (D) stimuli in the recall-set condition, then the combined or dual-access notion is suggested.

Method

Subjects.—Sixteen volunteer *S*s from the University of Washington served in Exp. IV, each participating in one hour-long experimental session. Four males and 4 females were assigned to the recognition-set condition; while 5 males and 3 females were run in the recall-set condition.

Stimuli, conditions, and procedure.—The two conditions of task expectation and the experimental procedure outlined for Exp. III were used in Exp. IV with three exceptions. First, during the recognition test, *S*s were asked to respond YES only to SS stimuli and NO to DS and D stimuli. Second, the 128 test stimuli presented during the RT recognition task consisted of 4 each of the 16 SS stimuli, 2 DS drawings for each of the 16 original stimuli, and all 4 drawings of 8 NEW stimulus objects—2 from each of the 4 semantic categories. The third difference concerns method of presentation for recognition test items: a slide projector with stimulus and response timing mechanisms paced the presentation of the pictures and measured RT to *S*'s responses.

Results

The NO responses were averaged over objects for each *S*. Since each different drawing appeared only once, no division by trials was made for Exp. IV RTs; the two resulting means, DS and D item RTs, were subtracted to form a difference score for each *S*. Analysis of variance contrasting groups and *S*s for the difference scores indicated that source groups was significant, $F\ (1, 14) = 8.93$, $p < .01$. While recognition-set *S*s did not differ in RT to DS and D items (mean RT = 891 and 895 msec., respectively), recall-set *S*s were slower to DS than to D items (mean RT = 964 and 892 msec., respectively). The results of the RT comparison are graphed in Fig. 4.

Errors occurred to 15.0% of the total items for the recognition-set condition and to 12.0% of the items for the recall-set group. False-alarm errors, those DS and D items incorrectly called "old," averaged 10.7% for recognition-set *S*s and 9.6% for recall-set *S*s. Analysis of variance of errors by item type failed to attain statistical significance for all effects. Mean DS and D errors for recognition-set *S*s were 10.2% and 11.3%; while mean DS and D errors for recall-set condition were 10.6% and 8.6%. Although group means do not reflect the error type differences found in Exp. III, it should be noted that only 3 of the 16 *S*s made more D than DS errors, and 2 of the 3 contributed almost half of the errors for the recognition-set condition. When error data for the 2 *S*s are ignored, mean errors for DS items are greater than D errors for both conditions: DS = 2.67 and D = 1.33 errors for recognition-set *S*s; and DS = 3.38 and D = 2.75 errors for recall-set *S*s.

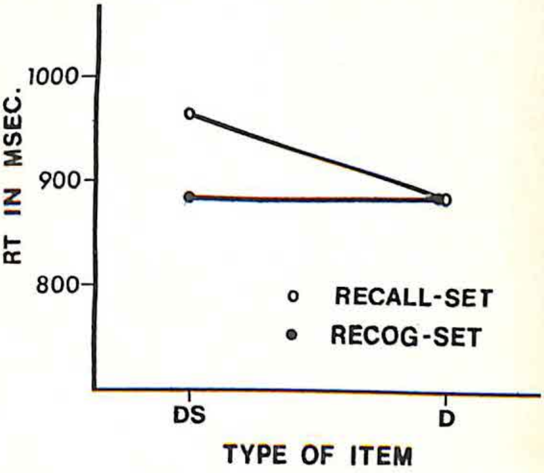


FIG. 4. Mean RT for NO responses in the stimulus-recognition task, Exp. IV.

The RTs for YES responses were not compared.

Discussion

Since recall-set *S*s make longer RTs to DS stimuli, we may conclude that they do not ignore name code memory in order to make the visual recognition match. Nevertheless, the fact that these *S*s make very few errors in the visual recognition task (Exp. IV) indicates that both name and orientation memory were affecting their judgments. Just as the double clustering indicates that recognition-set *S*s are storing and accessing both name and orientation, it also seems that recall-set encode and use both types of information during recognition.

Experiment III suggested that the combining of two memory codes during retrieval is more than a recall function; the combining appears for recognition-set *S*s in both recall and recognition paradigms. Experiment IV is consistent with a differential encoding explanation of the task expectation effects. Those *S*s set to expect a recognition task seem to access their memory for the pictorial stimuli through a visual code; while those *S*s expecting recall seem to access memory according to name information.

The differences in errors made during the recognition tasks (of Exp. III and IV) indicate that although the two groups differ in the modality through which memory is accessed (either visual or verbal), visual codes do affect the accuracy of recall-set *S*s' visual judgments and name codes do affect the accuracy of recognition-set *S*s' name judgments. Since the RT for these judgments do not differ (between DS and D NO judgments for recall-set *S*s and between SS and DS YES judgments for the recognition-set condition) a strong interpretation suggests that both codes are accessed together, regardless of the access modality. The error and RT data, taken together, certainly require a parallel access explanation, as suggested by Freedman and Loftus' (1971) "think of a word" task and by data from the letter-matching experiments reviewed by Posner and Keele (1970).

The visual and semantic influences on our memory for the things we see makes a complex picture, indeed. The experiments presented here are admittedly exploratory, but they suggest that very little of our "visual" memory can be separated from a verbal descriptive system or from an expectation of the way in which that memory will be used.

REFERENCES

BAHRICK, H. P., & BOUCHER, B. Retention of visual and verbal codes of the same stimuli. *Journal of Experimental Psychology*, 1968, **78**, 417–422.

BATTIG, W. F., & MONTAGUE, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 1969, **80** (3, Pt. 2).

BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 1953, **49**, 229–240.

BOUSFIELD, W. A., BERKOWITZ, H., & WHITMARSH, G. A. Associative clustering in the recall of minimally meaningful geometric designs. *Canadian Journal of Psychology*, 1959, **13**, 281–287.

BOUSFIELD, A. K., & BOUSFIELD, W. A. Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports*, 1966, 935–942.

BOWER, G. Organizational factors in memory. *Cognitive Psychology*, 1970, **1**, 18–46.

BYRNE, B., & BROOKS, L. R. Probing the organization of noun lists by inducing modality-specific conflict at output. Paper presented at the meeting of the Psychonomic Society, San Antonio, November 1970.

COFER, C. N., & BRUCE, D. R. Form-class as the basis for clustering in the recall of nonassociated words. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 386–389.

COHEN, B. H. An investigation of recoding in free recall. *Journal of Experimental Psychology*, 1963, **65**, 368–376.

COHEN, G. Search times for combinations of visual, phonemic, and semantic targets in reading prose. *Perception & Psychophysics*, 1970, **8**, 370–372.

DALRYMPLE-ALFORD, E. C. Measurement of clustering in free recall. *Psychological Bulletin*, 1970, **74**, 32–34.

ELLIS, H. Transfer of stimulus predifferentiation to shape recognition and identification learning: Role of properties of verbal labels. *Journal of Experimental Psychology*, 1968, **78**, 401–409.

FREEDMAN, J. L., & LOFTUS, E. F. Retrieval of words from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 107–115.

FROST, N. Clustering by visual shape in the free recall of pictorial stimuli. *Journal of Experimental Psychology*, 1971, **88**, 409–413.

FUNKHOUSER, G. R. Effects of differential encoding on recall. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 1016–1023.

MANDLER, G. Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. Vol. 1. New York: Academic Press, 1967.

MOSTELLER, F., & YOUTZ, C. Table of Freeman-Tukey arcsine transformation for binomial proportions. Unpublished table, Harvard University, 1954.

POSNER, M. I., & KEELE, S. W. Time and space as measures of mental operations. Paper presented at the meeting of the American Psychological Association, Miami, September 1970.

ROENKER, D. L., THOMPSON, C. P., & BROWN, S. C. Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, 1971, **76**, 45–48.

SAMPSON, J. R. Free recall of verbal and non-verbal stimuli. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 215–221.

SCOTT, K. G. Clustering with perceptual and symbolic stimuli in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 864–866.

SHEPARD, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 156–163.

SHUELL, T. J. Clustering and organization in free recall. *Psychological Bulletin*, 1969, **72**, 353–374.

STANDING, L., CONEZIO, J., & HABER, R. N. Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, 1970, **19**, 73–74.

TULVING, E. Subjective organization in free recall of "unrelated" words. *Psychological Review*, 1962, **69**, 344–354.

WICKELGREN, W. A. Multi-trace strength theory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

WRIGHT, P., & KAHNEMAN, D. Evidence for alternative strategies of sentence retention. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 197–213.

(Received January 3, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 327–333

U-SHAPED METACONTRAST FUNCTIONS WITH A DETECTION TASK[1]

SUE COX[2] AND WILLIAM N. DEMBER

University of Cincinnati

The present study was undertaken primarily to (*a*) determine whether U-shaped backward masking functions of detection in disc–ring metacontrast can be obtained with a forced-choice indicator and (*b*) assess the role of phenomenal brightening of the target in the shape of the masking function. Two luminance levels were employed. One group of *S*s was instructed to locate the black target disc within one of the two masking rings; another group was asked, in addition to the detection task, to report whether the target appeared black or white. Two luminance levels were employed with 11 values of interstimulus interval (ISI). The main effect of instructions failed to reach significance; effects of luminance and ISI as well as their interaction were significant at the .001 level. U-shaped functions were obtained. Analysis of the phenomenal reports led to the conclusion that the U minimum is a direct result of target indetectability, perhaps mediated by brightenings, rather than of *S*'s shifting criterion.

While **U**-shaped masking functions in visual backward masking are of great theoretical importance, their existence remains controversial for both discrimination and detection tasks. Studies employing discrimination of the position of a gap in the letter C combined with forced-choice procedures have yielded **U**-shaped relations between discrimination accuracy and the time interval separating target from mask (interstimulus interval, or ISI) (Stewart & Purcell, 1970; Weisstein, Jurkens, & Onderisin, 1970). **U** shapes have also been obtained with letters as target and ring as mask and letter identification as the response measure (Stewart & Purcell, 1970; Weisstein & Haber, 1965). However, employing similar stimuli and procedures, Eriksen has obtained considerable evidence that amount of masking is a monotonically decreasing function of ISI (Eriksen, 1966; Eriksen & Hoffman, 1963). The reasons for the discrepant findings are not at all obvious, especially since many of Eriksen's studies were attempts at direct replication. One possibility, however, may be the relatively low levels of luminance that Eriksen typically uses.

Although studies using judgments of apparent brightness have yielded **U** shapes (Alpern, 1953; Schiller & Smith, 1966), detection (reaction time) measures under the same experimental conditions reveal a constant function with no differences in reaction time to the masked target over a range of ISI values (Fehrer & Raab, 1962; Schiller & Smith, 1966). Some **U**-shaped masking functions have been obtained with detection as *S*'s task (Kolers, 1962; Kolers & Rosner, 1960; Werner, 1935), but they so far have been reported only when "nonobjective" response measures—that is, phenomenal reports—were employed. Eriksen and Marshall (1969), who used a forced-choice technique and a detection task, failed to replicate one of Kolers' **U**-shaped curves. This failure, coupled with the reaction time data, raise the possibility that **U**-shaped masking functions may only be found with discrimination tasks. The present study was undertaken primarily to determine whether a **U**-shaped function can be obtained with a detection measure as well. Furthermore, an "objective" (forced-choice) detection measure with disc as target and ring as mask was used.

It is important to establish the existence of **U**-shaped functions for theoretical reasons. Models of masking, such as the one proposed by Eriksen and Hoffman (1963), based solely on the notion of luminance summation–contrast reduction, cannot account for **U**-shapes. On the other hand, Weisstein's (1968) lateral inhibition model requires the occurrence of **U** shapes. The Purcell, Stewart, and Dember (1968) model is temporally neutral; that is, it makes no predictions regarding the temporal characteristics (monotonic or **U** shape) of masking functions. However, **U** shapes may result as artifacts of the target "brightenings" referred to in the Purcell et al. model. A second purpose of this study, then, was to examine the role of target brightenings in the shape of the masking function.

---

[1] Based on a master's thesis at the University of Cincinnati. Supported by Grant EY 00481-04 to William N. Dember from the National Eye Institute.

[2] Requests for reprints should be sent to Sue Cox, Department of Psychology, University of Cincinnati, Cincinnati, Ohio 45221.

According to the Purcell et al. (1968) model, when a target stimulus is presented (typically a black disc on a white background), the neurons serving the area of the disc fire at a lower rate than the surround neurons, which fire and subsequently inhibit themselves. When the masking ring follows, the disc neurons in the center of the ring, being less inhibited, fire at a much greater rate than the surround, resulting in a "brightness reversal." Through a temporal averaging of the black target presented, and the reversed white disc, the phenomenal appearance of the center of the ring can appear either darker than, the same as, or brighter than the background, depending on stimulus conditions. (Outcomes of the averaging process that are brighter than the background are called brightenings.) Conditions which facilitate reversals are those that serve to increase the initial differential inhibition, namely, increased target-field energy, dark fixation field and ISI interval, and brief ISIs. Thus, according to this model, Eriksen (1966; Eriksen & Hoffman, 1963; Eriksen & Marshall, 1969) may have failed to obtain **U** shapes because the luminance levels employed were too low.

The Purcell et al. (1968) model has been used to account for the recovery of a masked target through inhibition of the mask (Dember & Purcell, 1967) and has provided an interpretation of the phenomenal brightness reversal and re-reversal found in masking and recovery (Purcell & Dember, 1968). In addition, it can account for the decreasing spatial effectiveness of the mask with increasing ISI (Purcell et al., 1968) and for the decrease in masking found with increasing spatial separation between target and mask (Cox, Dember, & Sherrick, 1969). Even more important, the model had led to the surprising but verified prediction that an increase in its duration or luminance can render a target less detectable under masking conditions (Purcell, Stewart, & Dember, 1969).

Brightenings may be presumed to affect the shape of the masking function in the following manner. When the phenomenal appearance of the target is brighter than the surround, in a two-alternative forced-choice presentation, *S* is confronted with two rings, the center one of which is brighter (the one containing the disc) and the other, darker by comparison. Suppose the centers of the two rings are discriminable at all times but that *S* has difficulty deciding which of the two rings to respond to as containing the disc. Since *S*'s task is to "locate the black disc," he may choose the apparently darker but physically empty-centered ring. Sometimes *S* may decide that it is the brighter centered ring which contains the disc, perhaps on the basis of flicker due to the disc's presence. The *S* may shift between these two response criteria. Informal reports from *S*s suggest that this is typically what happens. But only one study has been reported which was explicitly designed to demonstrate the effect of brightenings on *S*'s performance in a forced-choice detection task (Heckenmueller & Dember, 1965). One group of *S*s was instructed, in the conventional manner, to locate the black disc; a second group was instructed to choose the ring whose interior looked most different from its background. The effect of this instructional difference was in the expected directions, but neither large nor at an impressive level of significance. However, the stimulus conditions were not optimal for obtaining **U** shapes in that the mask had more energy than the target.

In summary, then, the present experiment was designed primarily to investigate (*a*) the shape of the masking function in the context of a forced-choice detection task, where the parameters of stimulation were chosen to maximize the occurrence of **U** shapes, and (*b*) the possible role of target brightenings in determining the shape of the masking function. The design also allowed for examination of main and interactive effects of target-field luminance, a variable of special concern in the Purcell, et al. (1969) model.

Method

Subjects.—Eight male and eight female paid volunteers enrolled in psychology at the University of Cincinnati served as *S*s. All *S*s had normal or corrected-to-normal vision and had never participated in a masking experiment before. All *S*s were trained with all stimulus conditions for 2 days prior to the collection of the experimental data.

Stimuli.—The stimuli were black India-ink drawings on white index cards. The target consisted of a disc subtending 24′ of arc. The mask consisted of a pair of rings horizontally arranged; the inner diameter of each ring subtended 24′ of arc and the thickness of the ring walls subtended 12′. The separation between centers of the rings subtended 58′ of arc.

Apparatus.—The stimuli were presented tachistoscopically (Scientific Prototype GB). The target stimulus was attached to a plate which moved horizontally. The apparatus included adjustable stops so that the positions at which the disc was centered in each of the two rings could be marked and the disc easily shifted between these two positions. All *S*s were dark adapted for 6 min. before the presentation of the stimuli. Viewing was monocular with the right eye.

Procedure.—The stimuli were presented in the order: dark fixation field, target, dark ISI, mask, dark fixation field. The field of the tachistoscope used for the fixation field was also used for the ISI. It contained four peripheral red dots arranged in a diamond pattern to define the center of the field where the stimuli were presented. Two luminance conditions, 10 and 50 ftl., were employed, with the target luminance equal to that of the mask. The duration of both target and mask was 10 msec. Eleven ISI durations were used, ranging from 0 to 100 msec. in 10-msec. steps. Approximately 15 sec. elapsed between presentations.

Altogether there were 22 stimulus conditions (2 Luminance × 11 ISI). Each *S* received all 22 conditions in blocks of 10 trials per condition on each of the 5 experimental days. A trial consisted of positioning the disc in one of the two rings, signaling *S* to push a button to initiate the stimulus sequence, and *S*'s saying "right" or "left" to indicate in which of the two rings the disc appeared, guessing if necessary. The location of the disc on a given trial was random, as was the order of presentation of the conditions for a given *S* on a given day. On each of the experimental days, each *S* received 10 practice trials from a single condition chosen randomly for each *S* for each day before the collection of the data.

Half of the *S*s were randomly chosen to be in the Control group receiving the standard set of instructions according to the task, including: "Your task will be to tell me in which of the two rings the black disc appears. Sometimes it will be a difficult decision, but just try to do the best you can."

The remaining *S*s comprising the Experimental group received information about brightening along with the standard set of instructions:

> Your task will be to tell me in which of the two rings the disc appears. Sometimes, however, the ring which follows the disc makes the black disc appear dark and sometimes it may even appear to be light. Sometimes, it will be difficult to decide in which of the two rings the disc appears, but just try to do the best you can.
>
> In order to reduce errors, when you tell me whether the disc appears in the right- or left-hand ring also tell me whether the disc you saw was lighter or darker than the background.

The *S*s in both Experimental and Control groups received their instructions on the first day of practice and were reminded of them on each of the 6 remaining days of the experiment.

No *S* was given knowledge of results.

Results of Target Detectability

Percent correct detection corrected for chance guessing was calculated for each of the 16 *S*s for each of the 22 stimulus conditions. Means are presented in Table 1. Negative mean percent correct detection refers to performance below chance level. Mean percent detection for the Control and Experimental groups, summed across luminance and ISI, were 62.1 and 44.1, respectively; for low and high luminance, summed across groups and ISI, 71.5 and 34.8, respectively.

A three-way analysis of variance, with repeated measures on two factors, was computed on the data presented in Table 1. A summary of the results appears in Table 2. The groups main effect failed to reach significance, with the trend in the direction opposite that predicted. The effects of luminance and ISI, however, reached the .001 level of significance, as did the interaction between luminance and ISI. The

TABLE 1

Mean Percent Correct Disc Detection Corrected for Chance Guessing

Group	Luminance	Interstimulus interval (msec.)										
		0	10	20	30	40	50	60	70	80	90	100
C	Low	81.0	82.0	75.5	74.0	78.5	82.5	79.0	70.5	84.0	83.0	88.0
C	High	42.5	13.0	18.5	15.0	32.5	47.0	52.5	59.5	69.0	64.5	74.5
E	Low	72.0	66.5	60.0	44.5	64.5	59.5	60.5	61.5	65.5	76.5	64.0
E	High	16.5	5.0	−7.5	2.0	16.0	14.0	19.5	43.0	49.5	61.5	56.5

remaining F ratios obtained were less than 1.

The significant interaction between luminance and ISI was analyzed for trends to establish the existence of **U** shapes for the overall masking function and for low and high luminance functions separately. The low and high luminance functions depicted in Fig. 1 were tested separately for significant linear, quadratic, and cubic components by computing two additional analyses of variance. From these analyses were determined the total sum of squares and the error terms to be used in the trend analyses. The analysis of variance design used was the two-factor mixed design with repeated measures on one factor, the two factors being instruction condition and ISI. Table 3 summarizes the results obtained from the trend analyses.

Analysis of the low luminance curve as a function of ISI revealed a significant quadratic component ($p < .05$), and nonsignificant linear and cubic components. For the high luminance curve, the significant linear component establishes the rising portion of the curve, while the significant quadratic and cubic trends reveal the presence of a **U** shape. To further test the occurrence of **U** shapes, t tests were performed on the two curves presented in Fig. 1, using the appropriate error mean square from analyses of variance as the estimate of the variance. Significant differences in means at ISI = 0 and ISI = 30 were found for low luminance, $t\ (15) = 4.14$, two-tailed; and high luminance, $t\ (15) = 5.75$, both $p\text{s} < .001$.

TABLE 2

Analysis of Variance

Source	*df*	*MS*	*F*
Between *S*s	15		
Groups (A)	1	28,440.0	2.42
Error$_b$	14	11,752.8	
Within *S*s	336		
Luminance (B)	1	118,849.5	48.41*
ISI (C)	10	5,359.4	7.43*
A × B	1	158.3	.06
A × C	10	350.6	.49
B × C	10	3,137.2	11.12*
A × B × C	10	214.1	.76
Error 1	14	2,434.9	
Error 2	140	721.7	
Error 3	140	282.2	
Total	351		

* $p < .001$.

Discussion of Target Detectability

The difference between instruction conditions failed to reach significance, and the trend was in the opposite direction from the one expected. The failure of the Experimental group's instructions to increase disc detectability indicates that it may not be enough

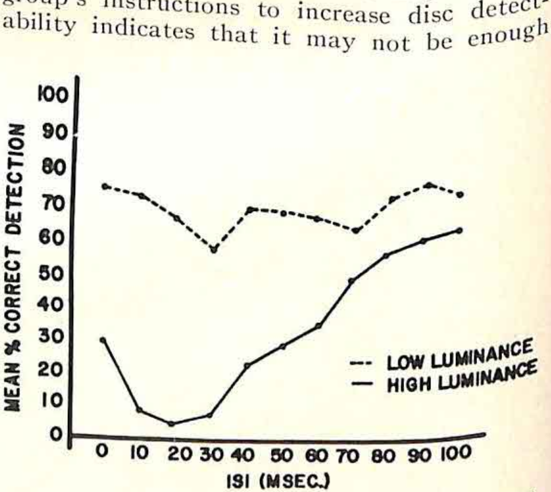


Fig. 1. Target detection as a function of ISI, disregarding groups.

simply to alert *S* that the disc may appear light, but that additional knowledge of when it will appear light may be required. If this were true, knowledge of results after each trial might have been more helpful. That the Control group tended to detect the disc better by about the same amount in all conditions probably indicates that the Experimental group's instructions merely made *S*'s task less clear to him.

Luminance definitely was a significant variable for disc detectability. When target-field luminance was relatively high, masking was facilitated. This is consistent with the finding by Purcell et al., (1969) that increasing target-field luminance leads to increased masking. Explanation of this effect according to the Purcell et al. (1968) model assumes that the greater the target field luminance, the greater the differential in neural inhibition between the region of the target disc and its illuminated target neurons and surround. When the mask is presented, the firing rate of the "target neurons" is greater than those of the surrounding field, making the target region appear white. The greater the differential inhibition, the whiter the target is made, and hence, the more completely its phenomenal appearance will be converted to a brightness equal to the background (because of the temporal averaging process of the original black target and the whitened target). It should be noted, of course, that in order to recognize the likelihood of finding U-shaped masking functions, target- and mask-field luminance were kept equal; thus, it is possible that in this study the luminance effect was carried at least in part by the mask field. A separate study to partial out target and mask effects is clearly called for.

Evidence for the existence of a U-shaped masking function was demonstrated by means of trend analysis and *t* tests, thus confirming the prediction of Weisstein's model (1968) that for target and mask of equal duration and luminance U-shaped functions result. The significant Luminance × ISI interaction suggests that the shape of masking function depends on the absolute luminance of target and mask, keeping ratios of target to mask luminance constant. This supports the suggestion that Eriksen's (1966) failures to obtain U shapes may result from the relatively low levels of luminance he uses, lower typically than the 10 ftl. used in the present study.

Inspection of the curves presented in Fig. 1 shows more masking in the 0- to 10-msec. ISI range than in the latter portions of the curves, making them appear more J than U shaped. It is possible that inclusion of some negative ISI values would have enabled the initial portion of the masking function to attain approximately the same height as the latter portion of the curve, thus completing the U. In any case, the critical issue is whether a linear or U shape better characterizes the obtained masking functions. Both the trend and the *t* tests indicate that the masking functions are better characterized as U shaped than as linear.

TABLE 3

Trend Analysis

ISI	*df*	*MS*	*F*
Low luminance			
	10	527	
Linear	1	397	1.45
Quadratic	1	1,324	4.84*
Cubic	1	413	1.51
Residual	7	447	
Error	140	274	
High luminance			
	10	7,970	
Linear	1	62,308	85.35***
Quadratic	1	6,415	8.79**
Cubic	1	9,105	12.47***
Residual	7	267	
Error	140	730	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Results and Discussion of Subjective Reports of Target Brightness

To test the hypothesis about the role of brightenings in the shape of the masking function, analysis of the subjective reports begins with organizing *S*'s responses into four categories: (*a*) disc correctly located and *S* reported that it appeared dark (correct dark); (*b*) disc correctly located and *S* reported it appeared light (correct light); (*c*) disc incorrectly located and appeared dark (incorrect dark); and (*d*) disc incorrectly located and appeared light (incorrect light). The percentage of each category of response presented separately for low and high luminance appears in Fig.

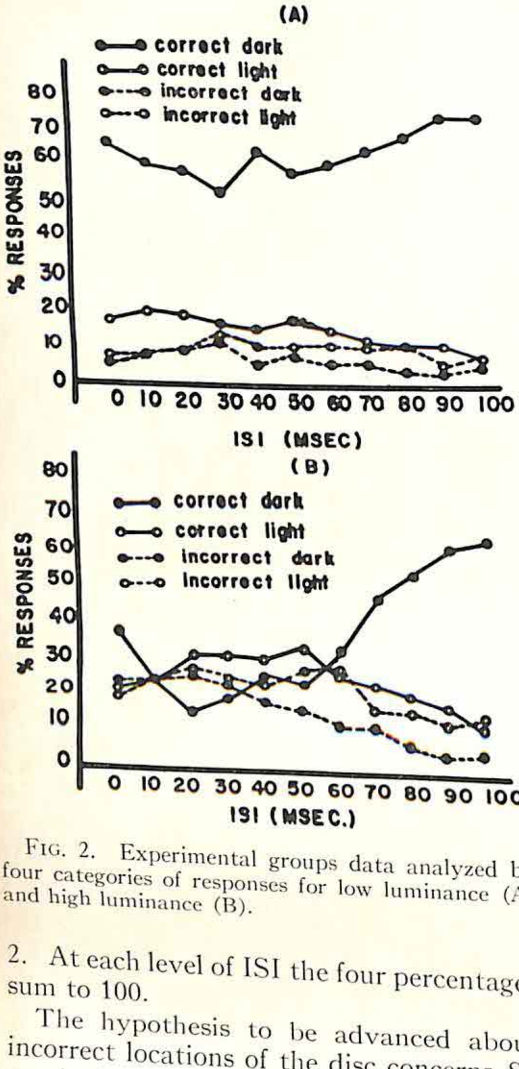


FIG. 2. Experimental groups data analyzed by four categories of responses for low luminance (A) and high luminance (B).

2. At each level of ISI the four percentages sum to 100.

The hypothesis to be advanced about incorrect locations of the disc concerns *S*'s employing an incorrect criterion. That is, suppose *S* can readily discriminate the centers of the two rings on the basis of brightness, one center appearing lighter or darker than the other. Further assume that *S* cannot decide whether the light- or the dark-centered ring contains the disc. Sometimes, then, the light-centered ring will be chosen as containing the disc, sometimes the dark-centered ring. It is expected, however, that for a given stimulus condition, the "correct" appearance of the disc is either one or the other. When *S* correctly locates the disc and it appears dark, for example, then when the disc is incorrectly located, it should appear light—assuming that the disc is at all times detectable and that incorrect locations of the disc are the result of *S*'s indecision about the "correct" appearance of the disc.

According to the above reasoning, general trends to be expected are the following: If the disc when correctly located appeared dark more often than light, incorrect locations of the disc for that condition should appear light more often than dark. Conversely, when the correctly located disc appeared light more often than dark, the incorrectly located disc should appear dark more often than light. Inspection of the relationships among the curves in Fig. 2 shows that the hypothesis is not tenable.

It may be concluded, then, that in addition to the instruction condition's failing to yield a difference in detection, the trends in the subjective data do not follow those expected if the "apparent" **U** minimum is a result not of target indetectability per se, but rather of *S*'s indecision as to the target's correct appearance. Quite the contrary, the minimum of the **U** seems to result directly from the target's indetectability, thus establishing the occurrence of the **U**-shaped masking function for at least one set of stimulus conditions via an objective detection measure. In addition, indetectability of the target may be mediated by brightenings in the manner suggested in the Purcell et al. (1968) model.

REFERENCES

ALPERN, M. Metacontrast. *Journal of the Optical Society of America*, 1953, **43**, 648–657.

COX, S. I., DEMBER, W. N., & SHERRICK, M. F. Effect or backward masking of spatial separation between target and mask contours and of target size. *Psychonomic Science*, 1969, **17**, 205–206.

DEMBER, W. N., & PURCELL, D. G. Recovery of masked visual targets by inhibition of the masking stimulus. *Science*, 1967, **157**, 1335–1336.

ERIKSEN, C. Temporal luminance summation effects in forward and backward masking. *Perception & Psychophysics*, 1966, **1**, 87–92.

ERIKSEN, C. W., & HOFFMAN, M. Form recognition at brief durations as a function of adapting field and interval between stimulations. *Journal of Experimental Psychology*, 1963, **66**, 485–499.

ERIKSEN, C. W., & MARSHALL, P. H. Failure to replicate a reported **U**-shaped visual masking function. *Psychonomic Science*, 1969, **15**, 195–196.

FEHRER, E., & RAAB, D. Reaction time to stimuli masked by metacontrast. *Journal of Experimental Psychology*, 1962, **63**, 143–147.

HECKENMUELLER, E. G., & DEMBER, W. N. Paradoxical brightening of a masked black disc. *Psychonomic Science*, 1965, **3**, 457–458.

KOLERS, P. Intensity and contour effects in visual masking. *Vision Research*, 1962, **2**, 277–294.

KOLERS, P., & ROSNER, B. On visual masking (metacontrast): Dichoptic observation. *American Journal of Psychology*, 1960, **73**, 2–21.

PURCELL, D. G., & DEMBER, W. N. The relation of phenomenal brightness reversal and re-reversal to backward masking and recovery. *Perception & Psychophysics*, 1968, **3**, 290–292.

PURCELL, D. G., STEWART, A. L., & DEMBER, W. N. Spatial effectiveness of the mask: Lateral inhibition in visual backward masking. *Perception & Psychophysics*, 1968, **4**, 344–346.

PURCELL, D. G., STEWART, A. L., & DEMBER, W. N. Backward masking: Facilitation through increased target-field luminance and duration. *Psychonomic Science*, 1969, **15**, 87–88.

SCHILLER, P. H., & SMITH, M. Detection in metacontrast. *Journal of Experimental Psychology*, 1966, **71**, 32–46.

STEWART, A. L., & PURCELL, D. G. **U**-shaped masking functions in visual backward masking: Effects of target configuration and retinal position. *Perception & Psychophysics*, 1970, **7**, 253–256.

WEISSTEIN, N. A Rashevsky-Landahl neural net: Simulation of metacontrast. *Psychological Review*, 1968, **75**, 494–521.

WEISSTEIN, N., & HABER, R. N. A **U**-shaped backward masking function in vision. *Psychonomic Science*, 1965, **2**, 75–76.

WEISSTEIN, N., JURKENS, T., & ONDERISIN, T. Effect of forced-choice vs. magnitude-estimation measures on the waveform of metacontrast functions. *Journal of the Optical Society of America*, 1970, **60**, 978–980.

WERNER, H. Studies on contour: I. Qualitative analysis. *American Journal of Psychology*, 1935, **47**, 40–64.

(Received January 3, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 334–336

DISPLAY SIZE AND THE DISTRIBUTION OF SEARCH TIMES[1]

IRA T. KAPLAN[2]
City College of the City University of New York
WILLIAM METLAY AND CLIFFORD T. LYONS
Hofstra University

The time required to find targets on a visual display was investigated by varying the number of targets and the total number of items on the display. The items were three-digit numbers, randomly arrayed, and the targets were those numbers whose digits summed to 14. Display size varied from 50 to 200 items, and the number of targets varied from 2 to 10. For each display condition, the cumulative distribution of search times was negatively accelerated and could be fitted by an exponential equation. The slope parameter of the fitted curve was independent of the number of targets, but was inversely proportional to the total number of items on the display. The shape of the distribution and its dependence on the number of display items were derived from a mathematical model. Application of the model to the data estimated the average scanning rate at 39 items/min. This estimate was confirmed in a control experiment which measured scanning rate directly.

Visual search and verbal recall tasks have in common the property that *S* searches for one or more target items located within a larger set of alternatives. Kaplan, Carvellas, and Metlay (1969) proposed a model to account for the exponential distribution of search times obtained when *S* searches for several targets either on a visual display or in a category of words stored in memory (cf. Indow & Togano, 1970). The model pertains to the case where *S* repeatedly selects one item at random from all the items on the display. The probability that each item selected is a target not found previously is $(T - F)/N$, where T is the number of targets on the display, F is the number of targets already found, and N is the total number of items. Thus, the rate of finding targets per unit time, t, is

$$\frac{dF}{dt} = \frac{k(T - F)}{N}, \qquad [1]$$

where k is the rate per unit time at which *S* selects and examines items. Integrating Equation 1 and substituting

$$p = \frac{k}{N} \qquad [2]$$

yields an exponential function for the cumulative distribution of search times:

$$F = T(1 - e^{-pt}). \qquad [3]$$

The parameter p determines the rate at which the number of targets found approaches T as an upper limit. This model has been used to describe the results of several studies of memory search (Kaplan et al., 1969; Kaplan, Carvellas, & Metlay 1971; Metlay, Handley, & Kaplan, 1971), but it has not yet been applied to the task of searching for several targets on a visual display. Although exponential distributions of visual search times have been reported (Krendel & Wodinsky, 1960; Williams, 1966), those distributions were obtained by presenting *S* with a succession of visual displays, each of which contained only one target.

The purpose of the present experiment was to test the search model by varying the number of targets and items on visual displays. The model predicts that the cumulative distribution of search times will be exponential in form, and that the slope parameter, p, will be independent of the number of targets, but will vary inversely with the total number of items per display.

---

[1] This study was supported by United States Public Health Service Grant EY-00384 from the National Eye Institute and by Hofstra University Grant 0580.

[2] Requests for reprints should be sent to Ira T. Kaplan, Department of Psychology, City College of the City University of New York, New York 10031.

METHOD

Subjects.—The *S*s were 10 students, 5 males and 5 females, in an experimental psychology course at Hofstra University.

Displays.—The displays were arrays of three-digit numbers printed by an IBM 1132 printer. Each display contained 50, 100, 150, or 200 three-digit numbers randomly located within a circular area in the center of an 11 × 15 in. page. The area of the circle was proportional to the number of items, so that the item density was constant at 3.5 three-digit numbers per square inch. The targets were numbers whose digits summed to 14, e.g., 734. The number of targets per display was 2, 4, 6, 8, or 10. Each *S* received nine different displays: four display sizes, 50, 100, 150, and 200 items all with 6 targets per display, plus four 100-item displays with 2, 4, 8, and 10 targets per display, and a practice display of 100 items with 6 targets. Ninety different displays were generated by an IBM 1130 computer, which randomly selected the numbers for each display.

Procedure.—The *S*s were tested as a group in a classroom. Every *S* received a stack of nine displays: the practice display on top, followed by the eight experimental displays in a different random order for each *S*. The instructions were to find all the numbers whose digits summed to 14 and to list them in a column on a sheet of paper. To mark time on his answer sheet, *S* drew a line under the last target he found, at a signal which was given every 30 sec. The displays were presented face down. At *E*'s signal, *S* turned the top display over and began immediately to search. After 5 min., *S* was told to stop searching, and the procedure was repeated with the next display.

RESULTS

The rate at which the cumulative response curves approached their limits was determined by the number of items on a display and was independent of the number of targets. Figure 1 shows the number of targets found as a function of time for five of the eight displays in this experiment. The curves drawn through the data points are graphs of Equation 3, fitted by an optimization procedure which minimized the sum of deviations squared, subject to the restrictions that *T* could not be greater than the number of targets on the display or less than the number of targets actually found. The upper limit approached by each curve is clearly determined by the number of targets on the display: 2, 6, or 10 targets. The three middle curves in the figure all approach the same limit, 6 targets, but the rate at which they rise is related to the number of items on the display: the 50-item curve rises fastest of the three, and the 150-item curve is the slowest.

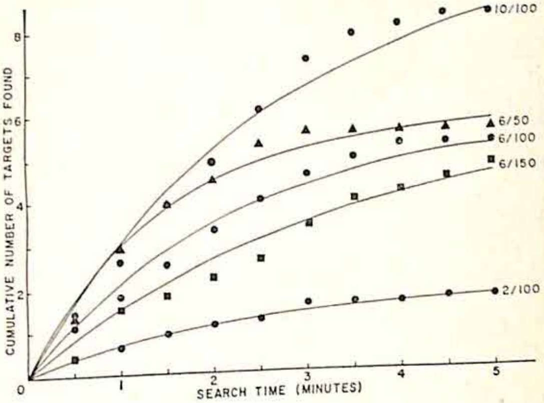


FIG. 1. Cumulative number of targets found as a function of time. (The label on each curve indicates the number of targets (2, 6, or 10) and the total number of items (50, 100, or 150) per display.)

The graphs for the three other displays have been omitted from Fig. 1 for the sake of clarity, but they are consistent with those shown. In every case, the estimated *T* values were equal or nearly equal to the number of targets on the display. The slope parameter p decreased as the number of items per display increased. The estimated p values were .69, .29, and .19 per minute for the 50-, 150-, and 200-item displays with 6 targets, and .40, .40, .44, .47, .37 per minute for the 100-item displays with 2, 4, 6, 8, and 10 targets, respectively. The p values for the 100-item displays show no consistent trend as a function of the number of targets.

It seems reasonable that the rate at which *S* scans a display should not vary appreciably with the number of items on the display. If the scanning rate, k, is in fact constant, then the model predicts that p should vary inversely with N. This prediction is supported by the results shown in Fig. 2, which is a plot of p versus N for all eight display conditions. The curve drawn through the data is a graph of Equation 2 with k held constant, fitted by the method of least squares. The value of

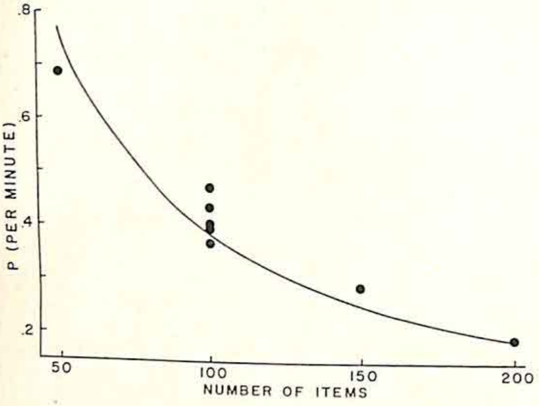


FIG. 2. The slope parameter p as a function of the number of items per display.

k estimated by this procedure was 395. items scanned per minute.

DISCUSSION

The results that the cumulative distribution of visual search times was exponential in form, and that the slope parameter, p, was inversely proportional to the number of items on the display confirm the predictions of the random search model. They are also consistent with the results of a memory-search study that varied the number of target words and the total number of items in a category stored in *S*'s memory (Metlay et al., 1971). Although the assertion that *S* selects targets entirely at random is a simplified description of his behavior, inspection of the data for individual *S*s revealed a distinctly unsystematic scanning pattern. The targets were not found in any consistent sequence, and the display was usually scanned two or three times before all the targets were found. In general, the cumulative distributions for individual *S*s exhibited the curvature characteristic of the group data, which indicates that the negative acceleration of the pooled distributions is not an artifact produced by averaging across *S*s.

Application of the search model to the data summarized in Fig. 2, implied that *S*s scanned the displays at an average rate of 39.5 three-digit numbers per minute. To test the validity of this estimate, an experiment was performed to measure *S*s' scanning rate more directly. Fifteen additional *S*s, also experimental psychology students at Hofstra, participated in the study. The task was to find, on a visual display that contained 100 three-digit numbers, those numbers whose digits summed to 14. The 100 numbers were listed in a column, in contrast to the principal experiment where the numbers were randomly arrayed, i.e., not in rows or columns. Six target numbers were located at preselected positions in the list. When *S* searched the list, he was instructed to scan from top to bottom, calling out the targets as he found them. His responses were tape recorded, and their latencies were transcribed from the tape with the aid of a stopwatch. The average scanning rate was determined by fitting a straight line by least squares to a plot of the average search time for each target against the position of the target in the list, according to the method used by Neisser (1963) and Kaplan and Carvellas (1965). The reciprocal of the slope estimated the scanning rate as 37.1 three-digit numbers per minute. This rate is in good agreement with the estimate of k obtained by fitting Equation 2 to the p values for the eight display conditions of the principal experiment, thus further confirming the predictive validity of the search model.

REFERENCES

INDOW, T., & TOGANO, K. On retrieving sequence from long-term memory. *Psychological Review*, 1970, **77**, 317–331.

KAPLAN, I. T., & CARVELLAS, T. Scanning for multiple targets. *Perceptual and Motor Skills*, 1965, **21**, 239–243.

KAPLAN, I. T., CARVELLAS, T., & METLAY, W. Searching for words in letter sets of varying size. *Journal of Experimental Psychology*, 1969, **82**, 377–380.

KAPLAN, I. T., CARVELLAS, T., & METLAY, W. Effect of context on verbal recall. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 207–212.

KRENDEL, W. S., & WODINSKY, J. Search in an unstructured visual field. *Journal of the Optical Society of America*, 1960, **50**, 562–568.

METLAY, W., HANDLEY, A., & KAPLAN, I. T. Memory search through categories of varying size. *Journal of Experimental Psychology*, 1971, **91**, 215–219.

NEISSER, U. Decision time without reaction time: Experiments in visual scanning. *American Journal of Psychology*, 1963, **76**, 376–385.

WILLIAMS, L. G. Target conspicuity and visual search. *Human Factors*, 1966, **8**, 89–92.

(Received January 8, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 337–342

RESPONSE LATENCY AS A FUNCTION OF HYPOTHESIS-TESTING STRATEGIES IN CONCEPT IDENTIFICATION[1]

RICHARD T. FINK[2]

University of New Mexico

The ability of M. Levine's subset-sampling assumptions to account for the decrease in response latency following the trial of the last error was investigated by employing a distributed stimulus set composed of four binary dimensions and a procedure which required *S*s to make an overt response in order to sample each dimension. Following Levine, *S*s were required to make an overt response indicating the trial on which they felt they had reached solution (ST). When the ST was used as a reference point it was found, in accord with Levine's subset-sampling assumption, that *S*s who sampled only one dimension per trial did not show the latency decrease following the trial of last error and that the latency decrease was due almost entirely to those *S*s entering the criterion run sampling more than one dimension.

In a recent article, Levine (1969b) presents evidence which questions the common assumption of early hypothesis-(H) testing models (Bower & Trabasso, 1964; Restle, 1962), viz., that *S* selects a single H, which determines his response, and feedback is evaluated in terms of the single H being tested. As an alternative, Levine (1969b) proposes that *S*s may sample a subset of Hs from a known finite H universe. From the subset, a working H which serves as a basis for *S*'s response is selected. The outcome of the response, whether "right" or "wrong" is used to evaluate Hs in the subset. If the response was correct, *S* maintains the working H and rejects those Hs in the subset which were not confirmed. If the response was incorrect, *S* rejects the working H and selects a new one from the confirmed Hs in the subset, or, if none of the Hs were confirmed, a new subset is established and a new working H selected. Levine (1969b) refers to this model of H sampling as the subset-sampling assumption (SSA). According to the SSA there are two important pivotal trials: the trial of last error (TLE), in which *S* first takes the correct H as his working H, and the solution trial (ST), in which *S* has eliminated the remainder of the subset and first holds only the correct H.

As a means of providing support for the SSA, Levine (1969a) measured latencies from the onset of the stimulus to *S*'s response from the TLE to the ST (indicated by *S* ringing a bell). It was assumed that latency is a direct function of the number of Hs being tested. As predicted, mean latencies decreased from TLE to ST and did not change from ST to the final criterion trial.

Although Levine's (1969a) predictions were confirmed, it is possible that the latency decrease between the TLE and ST was not due to a reduction in the number of Hs being evaluated. Several investigators (e.g., Millward, 1964) have reported, in paired-associate tasks employing more stimuli than responses, that when each *S*'s protocol for each item pair was aligned about the TLE that latencies decreased in a negatively decelerated fashion similar to that found by Levine (1969a). If latency decreases similar to those reported by Levine (1969a) can occur in paired-associate tasks in which *S*s reach solution by memorization instead of by H testing, then the reduction in latency found by Levine (1969a) could be due to something other than a reduction in the number of Hs being evaluated.

---

[1] This research is based on part of a dissertation submitted to the graduate school of the University of New Mexico in partial fulfillment of the requirements for the PhD degree and was supported in part by National Institute of Mental Health Grant HD03390. The author wished to thank Peder Johnson, who served as dissertation advisor and helped in the preparation of this article.

[2] Requests for reprints should be sent to Peder J. Johnson, Department of Psychology, University of New Mexico, Albuquerque, New Mexico 87106.

In Levine's (1969a) experiment, there was no means of directly assessing *S*'s H-sampling behavior since all values on all dimensions of the stimuli appeared together on each trial. The purpose of the current experiment was to provide a more direct test of Levine's (1969a, 1969b, 1970) SSA's description of H-sampling behavior by directly monitoring *S*'s dimensional sampling while recording post-TLE latencies. Operating under the assumption of Levine (Levine, 1969a, 1969b) and Trabasso and Bower (1968) that the dimension sampling or observing response is indicative of the H being tested, a procedure which required that *S*s make overt dimension-selection responses, was employed to permit the trial-by-trial monitoring of H sampling. The procedure involved the presentation of a distributed stimulus set composed of four binary dimensions. In order to observe the current value of any of the four dimensions, it was necessary for *S* to make a separate dimension-selection response to that dimension. Thus, it was possible to determine if the reduction in latency reflected the reduction in the number of Hs sampled.

METHOD

Subjects and design.—The *S*s were 52 volunteers from the introductory psychology classes at the University of New Mexico who participated in the experiment in order to obtain extra credit. Four *S*s were discarded from the experiment, 3 for failure to reach criterion on the first problem within 60 trials and 1 due to an apparatus malfunction. The remaining 48 *S*s were assigned to one of three groups of equal size. In each group, all *S*s recieved all eight unidimensional problems possible with four binary dimensions. The problem orders were represented by the use of an 8 × 8 Latin square in which problems had been randomly assigned to a letter. Each *S* was randomly allocated to a group and a row of the Latin square so that within each group there were two *S*s assigned to each row.

Apparatus and stimuli.—The apparatus involved a distributed stimulus set composed of four binary dimensions, each of which was represented by a column of 50 × 50 mm. squares. The current value of a dimension was indicated by back lighting one of the two squares in each column. A dimension selection panel consisting of four buttons was located directly below the stimulus display so that each of the four buttons was below one of the stimulus columns.

Response buttons and feedback lights were located on a separate panel below the dimension selection panel. The left response button and feedback light was labeled "yes," while the other pair was labeled "no." Feedback was administered by indicating the correct response with the appropriate feedback light. A stimulus-presentation button was located on the response panel between the two category response buttons. A response to the stimulus-presentation button presented the current values of the selected dimensions until the termination of feedback and made further dimension selection impossible. The ST button was located on a separate housing placed to the right of the stimulus display for *S*s to indicate when they felt that they had solved the problem. An enable light above the stimulus display was switched to green to signal the start of a trial and switched back to red concurrent with the offset of feedback.

Response latencies were recorded in hundredths of a second from a Hunter Klockounter. A Datatotal printout counter recorded the dimension-selection data and the response outcome.

Procedure.—Three groups were run in the experiment, two of which were permitted to select dimensions, while the remaining group (No Selection) was presented the current values on all four dimensions simultaneously on each trial. The *S*s in the Optional-Selection group could sample as many dimensions per trial as they wanted to, except on trials following the ST, where they could sample only the relevant dimension. In the other selection group (Single Selection), *S*s were permitted to sample only one dimension per trial throughout training.

Except where group differences necessitated changes, all *S*s were given the same instructions. The instructions began with a detailed description of the stimulus set using the third problem in each *S*'s series as an example during the instructions. When *S*s understood the stimulus display, they were told that they were to learn which stimuli were examples of the concept and which were not. They were further instructed that only one column, i.e., dimension, was relevant to solution and that they were to find that column. All *S*s were told that they would be given a series of problems where a problem could be repeated and that *E* would inform them when a new problem began. The *S*s in all groups were also told to press the ST button following feedback on the trial where they were certain that they had solved the problem. The Optional-Selection group was told that they could sample as many dimensions per trial as they wanted with the stipulation that following their ST response they sample only the one relevant dimension. The Single-Selection *S*s were instructed to locate the relevant dimension by sampling one dimension per trial. In the No-Selection group, the current values of all four dimensions were presented on all trials. In all three groups the stimulus values did not appear until a stimulus-presentation response was made.

A trial began when the enable light turned from red to green, which signaled *S*s in the dimension-selection groups to begin selecting dimensions and *S*s in the No-Selection group to press the stimulus-presentation button. The dimension selection *S*s

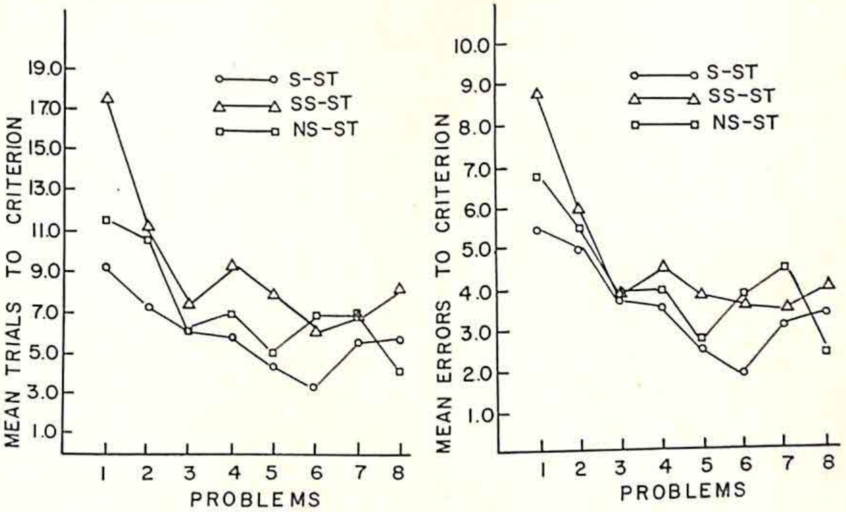


FIG. 1. Mean trials and errors to criterion for all groups and problems.

made the stimulus-presentation response when they were through selecting dimensions. Following the stimulus-presentation response, the stimuli remained on until the termination of feedback. Response latency was measured from the stimulus-presentation response to the category response. Immediately upon *S*'s category response, the feedback light above the correct response button was illuminated for 4.0 sec. Concurrent with the offset of feedback, the enable light changed back to red for a 3.0 -sec. duration until the start of the next trial. For all groups, a new problem was begun when *S* completed eight trials following his ST response.

RESULTS AND DISCUSSION

The trials- and errors-to-criterion data is presented in Fig. 1. The analysis of variance performed on the trials-to-criterion data showed that treatments, F (2, 45) = 5.89, $p < .01$; problems, F (7, 252) = 8.24, $p < .001$; and attributes/dimensions, F (4, 252) = 5.21, $p < .001$, effects all were significant. A similar analysis of the errors-to-criterion data found significant effects for problems, F (7, 252) = 9.13, and attribute/dimensions, F (4, 252) = 6.73, both $p < .001$. The treatments effects, found only with the trials data, was found to be due to a few *S*s in the Single-Selection group who inflated the trials data by testing all four dimensions before returning to the relevant dimension to make the ST response. The attribute/dimensions effect found in both the trials and errors analyses was due to a tendency for *S*s to formulate their H in terms of the top square of a column. Tukey's test indicated that the problems effect for both trials and errors data was due to the difference between the first problem and Problems 3–8 and between Problem 2 and Problems 5, 6, and 8. On the basis of the above analysis, the first two problems were treated as practice problems (Levine [1969a] also treated the first two problems as practice) and were excluded from further analysis. The remaining six were renumbered so that what was the third problem is now the first, etc., in subsequent analyses.

In the last six problems there were 16 occurrences of *S* making the ST response prior to reaching solution. This false ST response was ignored and the problem was continued following the normal procedure until a valid ST response occurred and criterion was achieved.

The data of primary importance involved response latency during the criterion run plotted about the ST. These data, including Levine's (1969a) experimental group, are presented in Fig. 2.[3] As can be seen

---

[3] In the present experiment, the ST was considered to be the trial immediately preceding the ST response. This differs from Levine (1969b) who considered the ST to be the trial following the ST response. Levine (personal communication, November 1969) later stated that since feedback on the trial prior to the ST is the critical event, that trial is actually the ST. Except when otherwise noted, Levine's data will be presented with the ST as defined here.

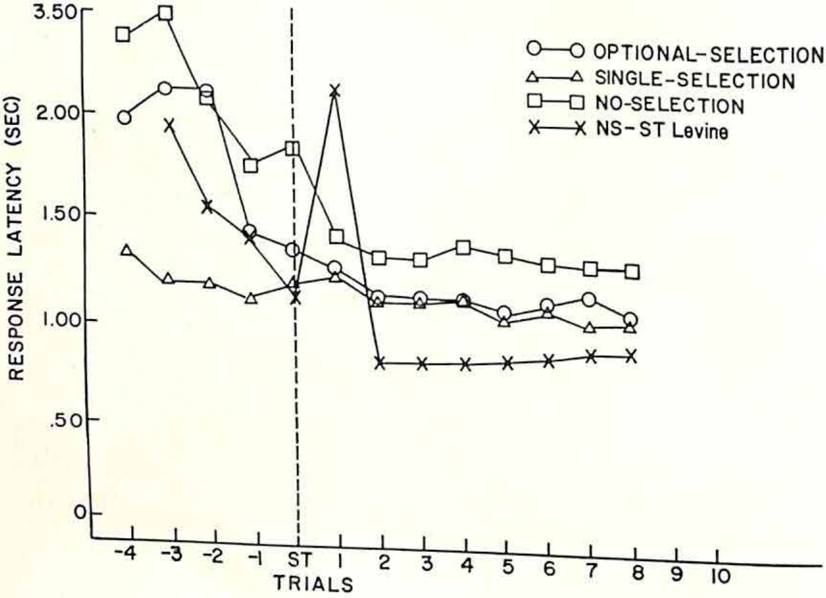


FIG. 2. Mean post-TLE response latencies plotted about the ST ($N > 32$ for all points).

from Fig. 2, the latency curves for the Optional-Selection and the No-Selection groups, in which *S*s were free to choose their H-testing strategy, closely parallel that of Levine's (1969a) experimental group, which was similar to the No-Selection group. The only major difference is the absence of the inflated latency on the first post-ST trial in the present experiment, which Levine attributed to a disruption caused by *S* making the ST response.

The absence of the positively decelerated latency decrease prior to the ST in the Single-Selection group supports Levine's (1969a) assumption that the response latency is a function of the number of Hs being evaluated. The pre-ST latency curve for the Single-Selection group should be flat, since this group could evaluate only one H per trial. However, inspection of Fig. 2 shows some evidence of a slight decrease in latency prior to the ST. It was not possible to perform a statistical analysis of the pre-ST portion of the latency data, since the variable number of trials between the TLE and ST caused the *N*s for the pre-ST trials to differ.

Since a fixed number of trials were given following the ST, it was possible to perform a series of analyses on the post-ST response latencies. For purposes of the analyses, the ST was considered to be the trial after the ST response, as was done by Levine 1969a. This definition of the ST was employed because of the possibility that the ST response caused the following trial to be inflated artificially, as was the case with Levine's (1969a) group.

An analysis of variance performed on the post-ST response latency data found significant effects for treatments, $F(2, 45) = 4.12$, $p < .05$; problems, $F(5, 252) = 4.38$, $p < .01$; and trials, $F(6, 270) = 2.28$, $p < .05$. Tukey's test showed that the treatments effect was due to the difference between Single-Selection and No-Selection groups. The problems effect was attributed to the difference of the first problem from Problems 3, 5, and 6 and to the difference between Problems 2 and 6. In order to determine more precisely the nature of the trials effect, a trend analysis was performed. The trials effect was found to have a significant linear component, $F(1, 45) = 8.1652$, $p < .01$; the remainder was not significant, $F(5, 225) < 1.0$. Both the linear component, $F(2, 45) < 1.0$, and the remainder, $F(10, 225) < 1.0$, were not significant for the Treatment × Trials interaction.

TABLE 1

SLOPES FOR POST-SOLUTION-TRIAL (ST) LATENCY

Exp. and group	Slope	95% confidence interval
Levine (1969a)		
Experimental	.020	
Present experiment		
Optional Selection	−.013	$-.091 \leq \mu \leq .065$
No Selection	−.008	$-.086 \leq \mu \leq .070$
Single Selection	−.022	$-.100 \leq \mu \leq .056$
$\bar{X}$ of all groups	−.014	$-.058 \leq \mu \leq .030$

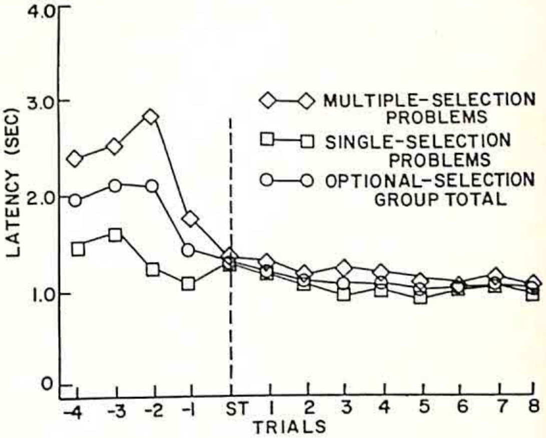


FIG. 3. Mean post-TLE response latencies plotted about the ST for multiple- and single-selection problems and for the Optional-Selection group total ($N > 22$ for all points).

The significant linear component in the trials effect presents a potentially serious discrepancy between Levine's (1969a) findings and those of the present experiment. Whereas the SSA predicts that the post-ST latencies should be constant and Levine (1969a) found a slight increase, a general post-ST decrease was found in the present experiment. In order to ascertain if this decrease constituted a significant deviation from the SSA, an analysis of the slopes of the line of best fit for the mean post-ST latencies for Levine's (1969a) group, for each of the three groups in the present experiment, and for the combined mean of all three groups, was performed.

The mean-square error term for the linear component from the trend analysis was used to compute the 95% confidence limits of the slopes of the lines of best fit for each individual group and the combined mean of all groups in the present experiment. As can be seen from Table 1, which presents these confidence intervals and the slopes of the lines of best fit, the zero slope predicted by the SSA falls within each of the confidence intervals, as does that for Levine's (1969a) group. To summarize, the results show a significant trials effect, but the slopes do not show a significant deviation from zero and therefore do not contradict Levine's SSA.

In order to ascertain whether the post-TLE latency decrease was typical only of *S*s entering the criterion run testing more than one H, the Optional-Selection group was partitioned into multiple-dimension-selection problems, in which *S*s sampled more than one dimension on any post-TLE trial and single-dimension-selection problems, in which only one dimension was sampled. Figure 3 shows the mean post-TLE response latencies plotted about the ST for single- and multiple-dimension-selection problems and the total for the Optional-Selection group. As can be seen, the multiple-selection problems account for the pre-ST latency decrease while the single-selection problems have relatively constant pre-ST response latencies, consistent with the SSA.

Additional evidence that the reduction in response latencies between the TLE and ST reflects a reduction in the number of Hs being evaluated is provided by Fig. 4,

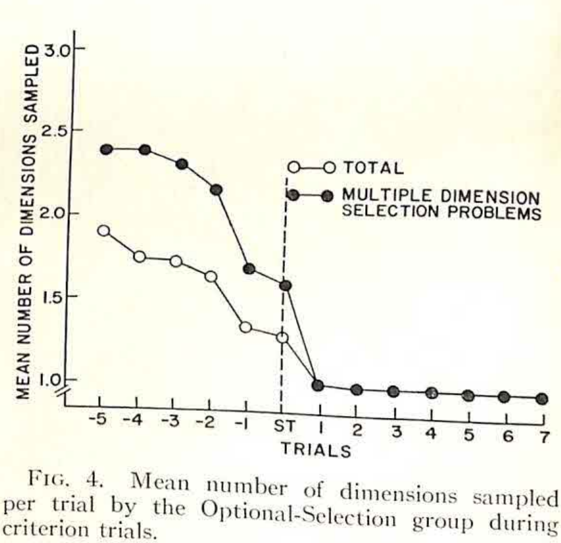


FIG. 4. Mean number of dimensions sampled per trial by the Optional-Selection group during criterion trials.

which presents the mean number of dimensions sampled per trial. As can be seen from Fig. 4, the number of dimensions sampled per trial decreases as the ST is approached reflecting the latency decrease and supporting the SSA.

In conclusion, it can be said that this experiment provided direct support of Levine's (1969a, 1969b) SSA. It was shown that some *S*s sample more than one dimension on at least one post-TLE trial and that the mean number of dimensions sampled per trial decreased between the TLE and the ST. Those problems where *S* sampled more than one dimension on any post-TLE trial were found to account for almost all of the post-TLE latency decrease prior to the ST.

REFERENCES

Bower, G. H., & Trabasso, T. R. Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology.* Stanford: Stanford University Press, 1964.

Levine, M. Latency-choice discrepancy in concept learning. *Journal of Experimental Psychology*, 1969, **82**, 1–4. (a)

Levine, M. Neo-noncontinuity theory. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 3. New York: Academic Press, 1969. (b)

Levine, M. Human discrimination learning: The subset-sampling assumption. *Psychological Bulletin*, 1970, **74**, 397–404.

Levine, M., Yoder, R. M., Kleinberg, J., & Rosenberg, J. The presolution paradox in discrimination learning. *Journal of Experimental Psychology*, 1968, **77**, 602–608.

Millward, R. Latency in a modified paired-associate learning experiment. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 309–316.

Restle, F. The selection of strategies in cue learning. *Psychological Review*, 1962, **69**, 329–343.

Trabasso, T. R., & Bower, G. H. *Attention in learning.* New York: Wiley, 1968.

(Received January 13, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 343–348

REPRESENTATION IN DEDUCTIVE PROBLEM SOLVING: THE MATRIX

STEVEN H. SCHWARTZ[1]
Wayne State University

DANIEL L. FATTALEH
University of Sidney

Seventy-two *S*s attempted to solve four "who-done-it" type deductive reasoning problems which varied in mode of representation, type of logical connective employed, and affirmative versus negative statement of information. Affirmative and conjunctive problems were solved more frequently than negative and disjunctive problems. However, the mode of representation in which the problem was presented had no effect. Analysis of the mode of representation actually employed (based on *S*s' written protocols) revealed a marked tendency for *S*s in various conditions to change the original representation of the problem. When these changes were taken cognizance of, the matrix representation was superior to other forms on all but negatively stated problems. The somewhat poorer performance of females on this task was related to their greater reluctance to change the representation of the problems.

In a previous study, Schwartz (1971), presented a number of "who-done-it" deductive reasoning problems in sentence form to 30 adult *S*s. An analysis of *S*'s protocols revealed a number of "modes of representation" by which *S*s organized the information presented in the problems. One form—the matrix—was clearly superior to other representations for achieving solutions to affirmatively stated problems. Success in these tasks appeared more closely linked to the ability to accurately represent information in some suitable format rather than the ability to draw logical conclusions from simple combinations of propositions. Task variables (such as number of relevant dimensions, logical connective employed, proportion affirmative instances) known to have considerable effect on the solution rates for traditional concept attainment tasks (Bourne, Ekstrand, & Dominowski, 1971) were effective to the extent to which they influenced the mode of representation *S*s employed. Only where differential use of the matrix representation occurred (in affirmative vs. negative problems) were different solution rates observed.

The present study attempted to replicate these findings within a more structured task. Mode of representation of information presented to *S* was controlled directly in order to determine more precisely the relationship between representation and problem-solving performance. It was hypothesized that *S*s presented problems in the matrix mode of representation would exhibit superior performance. The task problem was as follows:

Five convicts were standing in the prison yard talking.

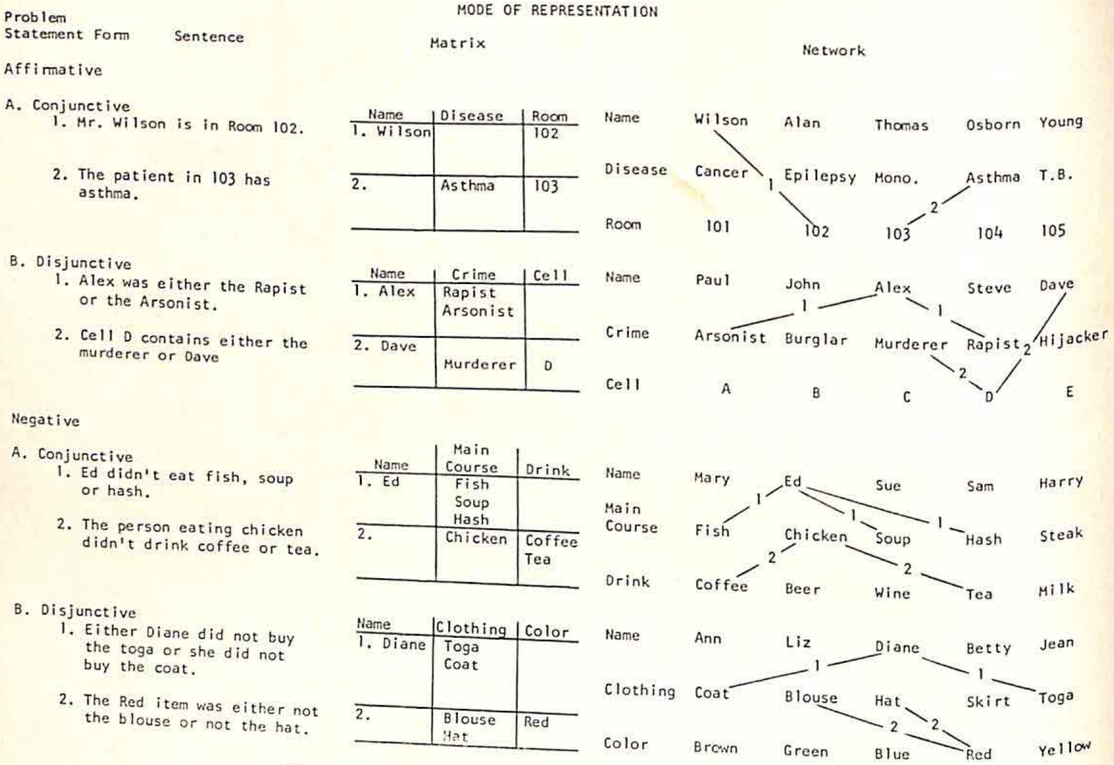
Each one was from a different cell block and each had committed a different crime.

1. Either Cell Block C did not contain Alex or it did not contain the burglar.
2. Either Cell Block A did not contain the murderer or it did not contain Zigowitz.
3. Either Young was not the embezzler or Young was not from Cell Block C.
4. One of the convicts was a murderer.
5. Either the hijacker was not Young or was not from Cell Block A.
6. Either the embezzler was not from Cell Block B or Zigowitz was not from Cell Block B.
7. Either the burglar was not from D or the burglar was not Thompson.
8. One of the convicts was from Cell Block E.
9. Smith was one of the convicts.
10. Thompson was the arsonist.

Who was in Cell Block E?

---

[1] Requests for reprints should be sent to Steven H. Schwartz, Department of Psychology, Wayne State University, Detroit, Michigan 48202.



Affirmative

A. Conjunctive

1. Mr. Wilson is in Room 102.
2. The patient in 103 has asthma.

Name	Disease	Room
1. Wilson		102
2.	Asthma	103

B. Disjunctive

1. Alex was either the Rapist or the Arsonist.
2. Cell D contains either the murderer or Dave

Name	Crime	Cell
1. Alex	Rapist Arsonist	
2. Dave	Murderer	D

Negative

A. Conjunctive

1. Ed didn't eat fish, soup or hash.
2. The person eating chicken didn't drink coffee or tea.

Name	Main Course	Drink
1. Ed	Fish Soup Hash	
2.	Chicken	Coffee Tea

B. Disjunctive

1. Either Diane did not buy the toga or she did not buy the coat.
2. The Red item was either not the blouse or not the hat.

Name	Clothing	Color
1. Diane	Toga Coat	
2.	Blouse Hat	Red

FIG. 1. Modes of representation presented to *S*s.

Please show all work on the accompanying work sheet.

METHOD

Seventy-two *S*s (36 males and females) were randomly drawn from the *S* pool at a large urban university and tested for about 2 hr. in groups of 12 to 15. All *S*s attempted to solve four "who-done-it" type deductive problems (Fig. 1), each containing information about five values on three dimensions (e.g., patients' name, room number, and illness). Two problems contained information presented in an affirmative fashion (e.g., the redhead bought the mini-skirt), the other two contained almost exclusively negative information (e.g., the maid was not in the bedroom). One problem in each of the above sets utilized conjunctive connectives (e.g., Ed had coffee and pie); the other, disjunctive connectives (Dave either wears a red or a blue tie). All problems contained sufficient information to determine all contingencies and contained an equal number of statements specifying various pairings. Some statements serve only to identify values (e.g., One of the men is John).

The design was a mixed factorial, with all *S*s attempting to solve an affirmative and negative, conjunctive and disjunctive problem (four problems in all).

Mode of representation served as a between-*S*s

TABLE 1

PERCENTAGE OF SUBJECTS ATTAINING SOLUTION UNDER EACH CONDITION

Mode of representation	Problem type						
	Conjunctive		Disjunctive		All problems male	All problems female	Totals all problems
	Affirmative	Negative	Affirmative	Negative			
Sentence	79	46	46	25	60	42	51
Matrix	83	50	46	25	54	46	50
Network	88	33	42	38	54	48	51
Totals	83	43	45	29	56	45	51

Note.—$n = 24$ per cell.

TABLE 2

Percentage of Subjects Changing Representation under each Condition

Mode of representation presented	Problem type						
	Conjunctive		Disjunctive		All problems male	All problems female	Totals all problems
	Affirmative	Negative	Affirmative	Negative			
Sentence	42	58	58	71	63	52	57
Matrix	8	8	21	29	17	17	17
Network	79	71	58	79	79	68	74
Totals	43	46	46	60	53	46	49

Note.—$n = 24$ per cell.

factor, with one-third of the *S*s receiving problems in a matrix format, one-third in a sentence format, and one-third in a network format (Fig. 1). (These modes had been observed in the previous study already mentioned).

In order to assure that *S*s understood the problem information in the mode presented to them, each *S* had to successfully complete a brief programmed instruction sequence. This required that he indicate in terms of equivalent English sentences the meaning of propositions presented in some mode of representation (Fig. 1). The *S*s were encouraged to show all work and had a maximum of 20 min. to attempt to solve each of the four problems. If he solved in less than the maximum allowed time, he went on to the next problem. The order of problems was randomized. After four problems, each *S* took a brief multiple-choice logic test which was designed to indicate the extent to which he could draw valid implications from *single* affirmative or negative, conjunctive and disjunctive statements.

Results

Effects of independent variable on solution rate.—The percentage of *S*s attaining solution under each condition is presented in Table 1. There are significant effects due to type of logical connective, F (1, 66) = 42.48, $p < .001$, and whether the information was presented in affirmative or negative form, F (1, 66) = 34.96, $p < .001$. However, contrary to expectations, mode of representation had virtually no effect on solution rate ($F < 1$). There was a tendency for males to do better than females, F (1, 66) = 4.85, $p < .05$, as well as a significant Logical Type of Problem × Affirmative–Negative Information interaction, F (1, 66) = 5. 44 $p < .05$. Although the affirmative problems were easier under all conditions, the advantage was greater on conjunctive as compared to disjunctive problems. No other effects were significant.

Analysis of representations actually used.—In view of the unexpected findings of no effects for mode of representation, the protocols were classified according to mode of representation *actually used* by each *S* in attempting to solve the problems. A classification scheme similar to that reported by Schwartz (1971) was employed. Tables 2 and 3 summarize the results of this analysis.

Table 2 indicates rather dramatically that *S*s tend to be active participants when attempting to solve these problems with almost half the problems changed in representation from the mode in which they were originally presented. Note the highly significant effect due to mode of representation, F (2, 66) = 22.12, $p < .001$, with only 17% of the *S*s in the matrix condition changing representation compared to 57% and 74% changes under the sentence and network treatments. Males tend to change representations somewhat more often than females 53% to 46%, F (1, 66) = 7.43, $p <$ .01. All other effects were not significant.

A more detailed analysis of these changes of representation and subsequent solution rates is presented in Table 3.[2] When mode of representation actually used is considered, the matrix representation does

---

[2] Significance tests are not reported in connection with the data in Table 3 since a particular subject may contribute anywhere from 0 to 4 entries to a cell, resulting in nonindependence of observations. The sample sizes are generally large, thus it is hoped the trends may be gleaned from the reported summary of the descriptive statistics.

TABLE 3

PROPORTION OF TIMES EACH MODE OF REPRESENTATION WAS USED AND CORRESPONDING SOLUTION RATES FOR VARIOUS PROBLEMS AND POPULATIONS

Mode of representation	Type of problem																			
	Affirmative (144 problem)		Negative (144 problem)		Conjunctive (144 problem)		Disjunctive (144 problem)		Matrix (96 problem)		Sentence (96 problem)		Network (96 problem)		Male (144 problem)		Female (144 problem)		Total (288 problem)	
	Use	Solution	Use	Solution	Use	Solution	Use	Solution	Use	Solution	Use	Solution	Use	Solution	Use	Solution	Use	Solution	Use	Solution
Matrix	.56	.80	.63	.37	.59	.71	.60	.45	.83	.53	.56	.59	.35	.68	.69	.64	.50	.51	.59	.58
Sentence	.26	.49	.22	.36	.30	.60	.18	.20	.17	.38	.39	.41	.19	.58	.24	.46	.24	.44	.24	.45
Network	.15	.43	.13	.22	.10	.42	.18	.28	.00	—	.00	—	.42	.31	.06	.33	.22	.30	.14	.31
Other	.03	.33	.02	.50	.01	.00	.03	.33	.00	—	.04	.25	.03	.00	.03	.50	.02	.33	.03	.35

yield better solution rates than the other modes of representation (58% compared to about 40%).

In comparing the use and solution rates for the matrix representation versus the three other modes combined, it was found that the matrix was used about equally often in affirmative and negative problems (56% and 63% respectively). The difference in solution rates between the matrix and other modes in the affirmative condition was large (80% vs. 46%) compared to those observed on negative problems (37% vs. 32%). Thus the advantage of the matrix representation appeared almost entirely in the affirmative problems.

There was little difference in use of matrix representation on conjunctive versus disjunctive problems with the matrix representation yielding superior performance on both types of problems (71% vs. 54% on conjunctive and 45% vs. 25% on disjunctive problems). Thus the matrix representation appeared about equally advantageous in both types of problems compared to the other representations.

As expected, the matrix was used more frequently in the condition where the problems were presented originally in that format (83%) compared to conditions where problems were presented in sentence (59%) or network formats (68%). Although the matrix representation produces higher solution rates under each condition, the relative superiority of the matrix over the other representations remains about the same in all three presentation conditions (with differences in solution rates of 15%, 19%, and 28%).

Males used the matrix representation 69% of the time compared to only 50% for females. In both cases, however, the matrix representation yields higher solution rates (64% vs. 44% for males, and 51% vs. 37% for females).

To summarize, Table 3 indicates superior performance for those problems attempted within the framework of a matrix representation in all conditions except for negatively stated problems.

Other transformations.—In addition to changes in mode of representation, two other transformations were determined from *S*'s protocols. In 45 problems, negative information was transformed into affirmative statements, compared to only 2 such cases in the opposite direction. Likewise, in 54 disjunctive problems information was changed into a conjunctive form, compared to only 17 conjunctive to disjunctive transformations. Performance in the logic posttest accounts for only 10% of the variance in success on the four experimental problems, r (71) = .34, $p < .01$. This is of the same magnitude as the .28 correlation between success on "who-done-it" problems and performance on this logic test found previously (Schwartz, 1971).

Discussion

The main findings of the previous study (Schwartz, 1971) are confirmed in that the matrix mode of representation, as actually used by *S*s, again leads to significantly better performance on all problems except those where information was presented in terms of negative statements. Furthermore, *S*s seem to recognize the appropriateness of this form of representation as evidenced in the low rate of changes in representation (17%) in the matrix as compared to the sentence and network conditions (57% and 74%). When *S*s do change the representation of the information, they most often change to a matrix representation; thus of the problems originally presented in sentence or network modes almost half (46%) were changed by *S*s into a matrix representation.

The generally superior performance on affirmative and conjunctive problems is in agreement with previous findings in the concept attainment literature (Conant & Trabasso, 1964; Schvaneldt, 1966). That the advantage of affirmative over negative problems is greater in the conjunctive compared to disjunctive case also corresponds to results found in studies on concept attainment (Bourne & Guy, 1968). Positive instances are more readily utilized in most strategies for attaining conjunctive concepts while the reverse is true for disjunctive concepts. The results, however, are contrary to the finding of little difference between performance in conjunctive and disjunctive problems reported by Schwartz (1971). The difference is primarily on disjunctive positive problems where solution rate was 77% in Schwartz (1971) compared to 45% in the present study. Scrutiny of the two disjunctive-affirmative problems (Schwartz, 1971) reveals that the statements in these problems inadvertently had been presented in an order conducive to solution. That is, information required to determine various attribute pairings appeared in close proximity to each other. Hence the conjunctive–disjunctive difference found in this study seems more reliable. However, as has been noted previously (Schwartz 1971), these results must be interpreted with caution since the problems were not precisely equated in terms of sentence-by-sentence information correspondence. Instead, since the primary objective of the study was to explore the utility of various modes of representation, the problems were matched only to the extent that each contained sufficient information for solution, an equal number of statements, and five specified values on each of three dimensions.

In addition, differences in solution rates could not be attributed solely to differential use of the matrix mode of representation. Rather, success seemed to be influenced more by how accurately information originally presented was represented by the *S*. For example, the poorer performance on negative problems may be inferred from the fact that 45 negative problems were transformed by *S*s into an affirmative mode while the reverse occurred only twice. This extra operation of transfering information provided additional opportunities for errors to take place. In fact there was evidence of incorrect representation of information on 29 of the 45 transformed negative problems. An intriguing ancillary finding in this study appears with respect to the small, but significantly superior performance of males (56%) compared to females (45%). Previous studies on problem solving with insight problems have reported similar results (Duncan, 1961; Staats, 1957). If, as hypothesized, development of an appropriate representation of information plays a key role in many problem-solving situations, then the tendency for females to be somewhat more reluctant than males (46%–53%) to change the problem from its original representation into a new mode may contribute to the small but consistent difference between performance of the sexes.

To conclude, the hypotheses of the instrumental role of representation of information in problem solving was supported. The superiority of a matrix mode of representation for a variety of problems has been demonstrated in two studies suggesting the need for subsequent investigation concerning why this form of representation is effective, and exploring tasks for which other representations may be more suitable.

REFERENCES

BOURNE, L. E. JR., EKSTRAND, B. R., & DOMINOWSKI, R. L. *The psychology of thinking.* Englewood Cliffs: Prentice-Hall, 1971.

BOURNE, L. E., JR., & GUY, D. E. Learning conceptual rules II. The role of positive and negative instances. *Journal of Experimental Psychology,* 1968, **77**, 488–494.

CONANT, M. D., & TRABASSO, T. Conjunctive and disjunctive concept formation under equal information conditions. *Journal of Experimental Psychology,* 1964, **57**, 250–255.

DUNCAN, C. P. Attempts to influence performance on an insight problem. *Psychological Reports,* 1961, **9**, 35–42.

SCHVANELDT, R. W. Concept identification as a function of probability of positive instances and number of relevant dimensions. *Journal of Experimental Psychology,* 1966, **72**, 649–660.

SCHWARTZ, S. H. Modes of representation and problem-solving: Well evolved is half solved. *Journal of Experimental Psychology,* 1971, **91**, 347–350.

STAATS, A. W. Verbal and instrumental response hierarchies and their relationship to problem solving. *Journal of Experimental Psychology,* 1957, **70**, 442–446.

(Received January 17, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 349–357

SEARCHING FOR TO-BE-FORGOTTEN MATERIAL IN A DIRECTED FORGETTING TASK

WILLIAM EPSTEIN[1] AND LUCINDA WILDER

University of Wisconsin

A selective search account of directed forgetting assumes that all input items are available in memory and that a cue to forget some of the items enhances recall of to-be-remembered (TBR) items by directing *S* to selectively search the TBR material and exclude the to-be-forgotten (TBF) material. Occasional tests of TBF items were included in a paired-associate probe experiment to test the notion that TBF items are available to *S* but are normally excluded from the search set. When *S* was not informed that the probe was for TBF material, recall was very low, as would be expected if the search were restricted to the TBR material. When *S* was informed that the probe was for TBF material and could redirect the search, recall was equal to lists in which no forget cue was given. The recall and error data provided good evidence that *S* can maintain a selective search. Selective search as the mechanism operative in directed forgetting was discussed.

When *S* is directed to forget part of the input, or to remember only part of the input, recall of the to-be-remembered (TBR) material is enhanced relative to recall of the same material when *S* is required to remember the entire input. This difference, called the Only effect, has been studied in paired-associate probe experiments by Bjork (1970, Exp. III), Shebilske, Wilder, and Epstein (1971), Epstein, Massaro, and Wilder (1972), as well as in other types of experiments (see Epstein, 1969, 1970; in press for a review of the literature). The prototypical trial in the paired-associate probe experiment proceeds as follows: the first set of pairs is presented, one at a time, then an interset interval, a second set of pairs is presented, followed immediately by a briefly exposed cue ("only first," "only second," or "either") that specifies *S*'s memory load for the trial, then a short retention interval culminating in a probe for one of the pairs. Superior recall on "only" trials relative to "either" trials defines the Only effect.

Epstein et al. (1972) attributed the Only effect to selective search. On "only" trials, the cue defines a search set and directs *S* to search selectively in that set. On "either" trials, the cue does not partition the memory set into search and not-search sets. Consequently, the search set on "only" trials is smaller and more likely to include the desired response than the search set on "either" trials. These differences between the search sets can account for the Only effect. Evidence for the selective search hypothesis was provided by Epstein et al. (1972) by showing that when the size and composition of the search sets are equated on "only" and "either" trials there is no advantage to "only" trials.

One implication of the selective search hypothesis which distinguishes it from its competitors, e.g., an erasure hypothesis or a selective rehearsal hypothesis (Epstein, in press), concerns the status of the to-be-forgotten (TBF) pairs. The selective search hypothesis implies that TBF and TBR items will be equally available. The results of recognition tests for TBF and TBR items, administered in other paradigms (e.g., Block, 1971; Elmes, Adams, & Roediger, 1970) support this claim. Nevertheless, the claim of equal availability does not imply equal recall. Recall test performance will depend on *S*'s strategy of search. Consider two conditions of testing for recall of a TBF item in the paired-associate probe paradigm. In both conditions, *S* is given an "only" cue, but then is tested unexpectedly with a probe stimulus from the excluded, TBF, set. In the informed condition, the stimulus is accompanied by a signal that indicates to *S* that the probe is for the TBF set; in the uninformed condition, the stimulus is presented alone. If *S* can maintain a selective search, performance on these two tests should differ significantly. In the informed condition, the signal should serve to direct *S* to search the noncued (TBF) set. Assuming availability of the TBF items, recall should be high. Ideally, recall of the TBF items should be superior to recall on an "either" trial and approximately equal to TBR recall on an "only" trial. Reitman, Malin, Tanner, Bjork, and Higman (1971) have shown that TBF probes that are identified for *S* do yield evidence of significant TBF recall. Their paradigm allowed selective rehearsal of TBR items, and this may account for their finding that TBF recall was consistently lower than TBR recall. The outcome on a TBF probe in the uninformed condition should be drastically different. Under these circumstances, *S* should restrict his search to the cued, but now inappropriate, set. Consequently, *S* should fail to locate the correct response and TBF recall should be poorer than recall on an "either" trial. The principal aim of the present experiments was to compare TBF recall under the informed and uninformed conditions with recall of TBR items on straight "only" and "either" trials.

---

[1] Requests for reprints should be sent to William Epstein, Department of Psychology, University of Wisconsin, Psychology Building, Charter at Johnson, Madison, Wisconsin 53206.

A secondary aim of the experiment was to reexamine our earlier findings regarding the absence of an influence of the opportunity for selective rehearsal during the postcue interval on the effectiveness of the cues. Both in the Shebilske et al. (1971) and Epstein et al. (1972) studies, the opportunity for rehearsal was varied between *S*s. The studies of incentives in acquisition and retrieval (Harley, 1965a, 1965b, 1968; Loftus & Wickens, 1970) show that the optimal conditions for demonstrating an incentive effect are those which expose the same *S* to all incentive conditions. The view that incentive manipulations essentially are rehearsal manipulations (Montague, Kiess, Hillix & Harris, 1970; Wickens & Simpson, 1968) suggested to us that the within-*S* design might provide a fairer test of the effect of rehearsal in the directed forgetting experiment. As in the earlier studies, the interesting comparison is between the Only effect when rehearsal is encouraged and when it is inhibited by a distractor task that fills the postcue interval.

METHOD

Apparatus.—All materials were presented on slides by means of a Kodak Carousel projector. A Psionix Series 1600 logic system controlled the presentation rate.

Design.—Forty-four lists (4 practice, 16 training, and 24 experimental lists) containing four pairs were presented to each *S*. The stimuli were CVC nonsense syllables selected from the Archer (1960) norms (range = 37 to 61); the responses were chosen from the categorized lists of common English words prepared by Battig and Montague (1969). All lists were categorized; i.e., the words serving as responses in the two sets in each list were selected from different categories, for example, BAF-DOOR, TEV-WALL, and CEB-SPIDER, QYT-ROACH. One pair in each list was arbitrarily chosen as the test item. The lists were presented as follows: after a 1-sec. ready signal two pairs were presented one at a time, a subtraction problem appeared for 3 sec., two more pairs were presented one at a time, followed by a .10-sec. exposure of the instruction cue, a 2-sec. postcue interval, and finally a test slide. The pairs were presented for 2 sec. each, the test slide was presented for 10 sec., and a 2-sec. intertrial interval preceded the next ready signal.

Instruction cue, serial position of the test item, and composition of the postcue interval were three within-*S* variables. For a given list the instruction cue was the word "first," "second," or "either." "First" and "second" were "only" cues in that they informed *S* that the tested pair would only come from the designated set, so that the other set could be forgotten. The cue "either" informed *S* that the tested pair would be selected from either the first or second set so that all pairs had to be retained. In the 16 training lists each of the four serial positions was tested once with an "only" cue and once with an "either" cue. Each of the eight cue-position combinations was tested once in a list in which the postcue interval was unfilled and once in a list in which the interval was filled with a subtraction problem.

The 24 experimental lists included tests of these 16 cue-position-filler combinations. In the remaining 8 lists, an "only" cue was presented, but the test item was a member of the TBF set rather than the TBR set. At the time of the test, *S* was informed or not informed that the cue had been inaccurate, i.e., that the test item was a member of the TBF set. On informed TBF (ITBF) probes, the stimulus member of the tested pair was accompanied by an asterisk. On uninformed TBF (UTBF) probes, the stimulus alone was presented. Four of the lists contained ITBF tests and four contained UTBF tests. Each type of TBF test was given once after the cue "first" and once after the cue "second" when the postcue interval was filled and unfilled. For half of the *S*s, tested TBF items had appeared in Serial Positions 1 and 3; for the remaining *S*s, tested TBF items had appeared in Positions 2 and 4.

The test pairs were counterbalanced across *S*s so that in the training lists and in the experimental lists each test item was tested in every cue-position-filler combination. This required 48 versions of the basic set of lists and two *S*s were assigned to each version. The presentation sequence of each version was random with the restriction that there were no runs greater than two of the same cue and same serial position of the test item. There was no restriction on the length of a sequence of filled and unfilled postcue intervals. In the experimental lists the eight TBF probes occurred between the fifth and twenty-first lists of the session.

All *S*s received the same set of four practice lists in the same order. In two of the lists the postcue interval was filled and it was unfilled in the remainder. The set contained three "only" cues and one "either" cue.

Procedure.—First, *S* was trained in the directed forgetting task by having *S* work through the 4 practice and 16 training lists. Then, instructions describing the occasional tests of TBF items were introduced and *S* was given the 24 experimental lists. The session was concluded with a questionnaire and short interview.

The tape-recorded instructions before the practice and training lists were similar to those used by Shebilske et al. (1971) and Epstein et al. (1972). The meaning of the cues was carefully explained and it was emphasized that the cue would always correctly represent the test. In the present experiment, however, *S* was specifically instructed to "*forget* the set that will not be tested" rather than told "you can forget the set that will not be tested" as was the practice in our previous studies. The *S*s were told that the composition of the postcue interval had been randomly determined and that runs of filled or unfilled intervals need not concern them. Whenever a subtraction problem appeared (in the middle of the list or in the postcue interval), *S*s were instructed to write the answer to the problem and classify the answer as odd or even. They were instructed to confine work on the problem to the 3-sec. interval that the subtraction slide was on the screen. When the postcue interval was unfilled, *S*s were instructed to use this time to rehearse the set(s) of pairs on which they would be tested. The *S*s were informed that the lists were categorized and that each pair would be presented only once during the experiment. The *S*s wrote their answers to the subtraction problems and the probe tests in a booklet. They were encouraged to write a response for each test. After each response, *S*s turned the page so that previous responses were not visible when a new list was presented. The *S*s were not allowed to change or make additions to previous responses.

After completion of the training trials, *S* was given instructions concerning the introduction of TBF probes. The purpose of these instructions was to alert *S* to the occasional ITBF tests, but to encourage maintenance of the same performance strategy that had been used during the training trials. The *S* was told that she would be able to recognize the TBF probes because the test syllable would be accompanied by an asterisk. It was stressed that these TBF tests would be very infrequent and that *S* should perform as if the cues were always accurate. The *S*s were told that "the purpose of these occasional tests of TBF items is to let us compare what happens to pairs you are trying to remember with pairs you are trying to forget. We won't be able to do this unless you continue trying to forget items you are told to forget." The *S*s were instructed to do as well as they could on TBF tests. It was again emphasized that the "most important thing is to continue doing the task the same way you have up until now." The *S*s were not told that the experimental trials would contain four UTBF probes. Finally, *E* informed *S* that all other aspects of the task, i.e., list composition, presentation rate, and method of making responses, were the same as they had been during the training lists.

The experiment was concluded with a questionnaire. The main purpose of the questionnaire was to determine if *S* felt she had changed her performance strategy after introduction of the TBF tests. The *S* was asked if she tried to do the task in the same way and if she felt she had succeeded in that goal. In addition, *S* answered several questions about how she had tried to learn the pairs, how she had utilized the cues, and whether she had rehearsed during the postcue interval. The interview was used to resolve ambiguities in *S*'s written answers.

Our predictions for the experimental trials were based on the assumption that *S*s would continue to utilize the "only" and "either" cues on the experimental trials as they had on the training trials. Therefore, only *S*s who had claimed not to have changed their performance strategy after the introduction of TBF probes were included in our analysis. When *S* claimed to have changed strategy, the data were discarded and a replacement *S* was given the same sequence of lists. The most common type of strategy change in the experimental trials involved trying to remember all the pairs in each list regardless of which cue was given.

Subjects.—The *S*s were volunteers from the university community who were paid $2.00 for their services; 107 *S*s were run to obtain the required 96 *S*s.

RESULTS

Only effect on training and experimental trials.—The purpose of this analysis was to ascertain whether introduction of TBF tests changed *S*'s performance on "only" and "either" trials. Recall on the 16 training trials was compared with recall on the 16 experimental trials on which an "only" cue or an "either" cue was given. Each *S* contributed data for all four serial positions. Overall performance on the training and experimental trials did not differ, F (1, 95) = 1.24, $p > .05$; nor was this variable (training vs. experimental trials) involved in any interactions. This analysis establishes that introduction of TBF probes did not alter performance on "only" and "either" trials. Therefore, performance on all 32 "only" and "either" trials may be described together.

The proportion recalled is shown in Fig. 1 as a function of cue and serial position for the filled and unfilled postcue intervals. Proportion recalled was significantly higher on "only" trials (.68) than on "either" trials (.53), F (1, 95) = 70.00, $p < .01$, and was higher following an unfilled postcue interval (.64) than following a filled postcue interval (.57), F (1, 95) = 24.65, $p < .01$. The Cue × Filler interaction was significant, F (1, 95) = 7.63 $p < .01$. Planned comparison tests showed that there was a significant 19% Only effect when the postcue interval was unfilled, F (1, 95) = 72.20, and a significant 10% Only effect when the interval was filled, F (1, 95) = 20.40, both $ps < .01$. The larger Only effect with unfilled intervals was due to significantly greater "only" recall for unfilled intervals (.74) than for filled intervals (.62) F (1, 95) = 29.28, $p < .01$. "Either" recall in the unfilled (.55) and filled (.52) conditions did not differ, F (1, 95) = 2.05, $p > .05$.

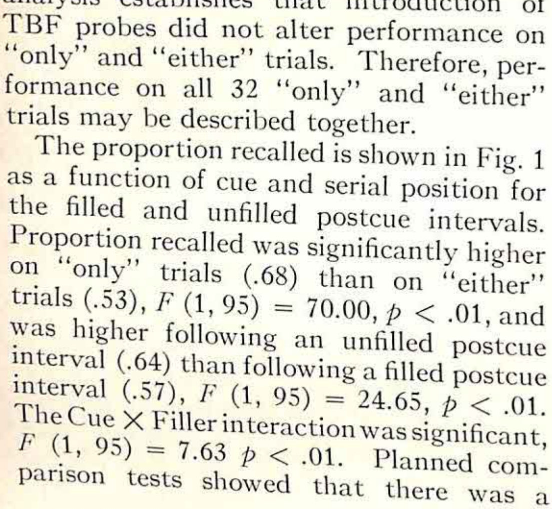


FIG. 1. Proportion correct for "only" cue and "either" cue for each of the four serial positions averaged over training and experimental trials. (Left panel shows results for filled condition; right panel, for unfilled condition. Data are from 96 original *S*s.)

Performance on experimental trials.—The analysis of performance on the experimental trials compared recall on "only," "either," ITBF, and UTBF probes. Since each S was given ITBF and UTBF tests on items that had appeared in two of the four serial positions in the list, only the data from the same two serial positons on "only" and "either" trials were considered in this analysis. For half of the *S*s, recall data for the four types of tests were generated by items that had appeared in the first and third serial positions; for the remaining *S*s, recall data for the four types of tests stemmed from the second and fourth serial positions.

Serial positions tested (one and three or two and four) was a between-*S* variable; it was found to be nonsignificant, F (1, 94) < 1, and did not enter into any interactions. Proportion recalled was significantly higher when the postcue interval was unfilled (.53) than when the interval was filled (.48), F (1, 94) = 6.05, $p < .05$. The main effect of cue was significant, F (3, 282) = 69.70, $p < .01$; proportions recalled were .70 on "only" trials, .58 on "either" trials, .51 on ITBF trials, and .22 on UTBF trials. However, there was a significant Cue × Filler interaction, F (3, 282) = 4.04, $p < .01$, as shown in Table 1. When the postcue interval was filled, the 7% Only effect was not significant, F (1, 282) = 2.66, $p > .05$. The ITBF recall was significantly lower than "either" recall, F (1, 282) = 7.81, and UTBF recall was significantly lower than ITBF recall, F (1, 282) = 18.62, both $ps < .01$. When the postcue interval was unfilled, the 16% Only effect was significant, F (1, 282) = 13.77, $p < .01$. The ITBF recall did

TABLE 1

Proportion Correct on Experimental Trials for Two Subject Samples

S and cue cond.	Postcue interval	
	Filled	Unfilled
Original		
Only	.64	.76
Either	.57	.60
ITBF	.44	.58
UTBF	.25	.19
Criterial		
Only	.72	.83
Either	.45	.50
ITBF	.41	.46
UTBF	.18	.13

Note.—$N = 96$ for original and criterial *S*s.

not differ from "either" recall, $F\ (1, 282) < 1$, but UTBF recall was significantly lower than ITBF recall, $F\ (1, 282) = 76.05$, $p < .01$.

Application of a performance criterion.—It will be remembered that the analysis of performance on training and experimental trials which was based on data from all four serial positions revealed no significant change in the Only effect after introduction of TBF tests. However, the analysis of experimental trials which was based on data from two serial positions indicated that the 7% Only effect in the filled condition was not significant. Failure to find a significant Only effect suggested that *S*s may have changed their performance for the positions in question. If *S*s did not change their performance strategy, the size of the Only effect for each *S* should be the same before and after introduction of TBF tests. We therefore devised a performance criterion indicative of *S*'s ability to take advantage of the "only" cue on training and experimental trials. We determined the number of *S*s who, after introduction of TBF tests, produced an Only effect that was as large or larger than the Only effect produced before introduction of TBF tests. In this manner, we segregated those *S*s who were equally able or more able to use the "only" cue after introduction of the TBF tests. It should be noted that this group did not consist solely of *S*s who produced a positive Only effect on the experimental trials; it also included *S*s who produced no Only effect or a negative Only effect. We found that 49 of the original 96 *S*s met the performance criterion; of these, 25 had been tested in Positions 0 and 3 and 24 had been tested in Positions 2 and 4.

An analysis of recall on experimental trials for these 49 criterial *S*s indicated that there was a significant Only effect in both the filled and unfilled conditions. A significant Only effect is a necessary prerequisite for interpreting ITBF and UTBF recall because the presence of an Only effect demonstrates that *S*s continue to take advantage of the forget instruction even though they are aware of occasional TBF probes. However, because of the small number of criterial *S*s, test items were not counterbalanced over the cue-position combinations as they had been in the original design requiring 96 *S*s. We therefore decided to replace the 47 original *S*s who had failed to meet the performance criterion.

Replacement Ss.—The data from the 49 *S*s who had met the verbal report and performance criteria were retained. Each of the 47 *S*s who had failed to meet the performance criterion was replaced by an *S* whose verbal report and recall data indicated she had used the same performance strategy on training and experimental trials. Replacement *S*s were run on each of the sets of lists requiring replacement until one was obtained that met the verbal report and performance criteria. A total of 126 *S*s were run to obtain the 47 required replacement *S*s; 27 of the 126 *S*s were disqualified by verbal report of a change in strategy and 52 *S*s were disqualified by failure to meet the performance criterion. The *S*s were undergraduate students who participated in the experiment to satisfy a course requirement.

The only effect on training and experimental trials for the 96 criterial Ss.—The proportion recalled on training and experimental trials is shown in Fig. 2 as a function of cue and serial position for the filled and unfilled postcue intervals. Proportion recalled on "only" trials (.70) was significantly greater than on "either"

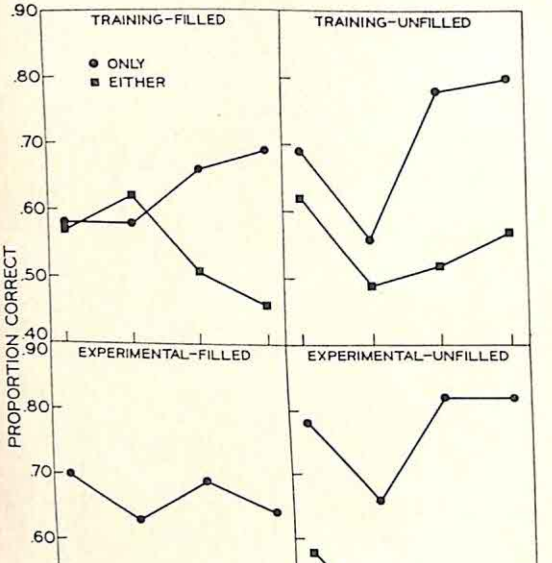


FIG. 2. Proportion correct for each serial position on training and experimental trials for "only" and "either" cues with filled and unfilled postcue intervals. (Data are from 96 criterial *S*s.)

trials (.52), F (1, 95) = 102.09, $p < .01$. Overall performance on training and experimental trials did not differ, F (1, 95) < 1. However, there was a significant Cue × Trials interaction, F (1, 95) = 13.93, $p < .01$. There was a significant 12% Only effect on the training trials, F (1, 95) = 36.60, and a significant 22% Only effect on experimental trials, F (1, 95) = 122.10, both ps $< .01$. The larger Only effect on experimental trials was due to greater "only" recall on experimental trials (.72) than on training trials (.67), F (1, 95) = 6.00, $p < .05$, and to smaller "either" recall on experimental trials (.50) than on training trials (.55), F (1, 95) = 6.50, $p < .05$.

Performance was significantly lower when the postcue interval was filled (.58) than when it was unfilled (.64), F (1, 95) = 12.95, $p < .01$. The Cue × Filler interaction was of borderline significance, F (1, 95) = 3.74, $p < .10$. However, planned comparison tests revealed that for the training trials the 16% Only effect in the unfilled condition was significantly larger than the 9% Only effect in the filled condition, F (1, 95) = 4.90, $p < .05$. For the experimental trials the difference between the Only effect in the filled and unfilled conditions (25% and 19%, respectively) fell short of statistical significance, F (1, 95) = 3.97.

Performance on experimental trials for 96 criterial Ss.—Serial position tested (1 and 3 or 2 and 4) was a between-*S* variable; it was found to be nonsignificant, F (1, 94) < 1, and did not enter into any interactions. The main effect of cue was significant, F (3, 282) = 126.23, $p < .01$; proportions recalled were .78 on "only" trials, .47 on "either" trials, .44 on ITBF trials, and .15 on UTBF trials. In this analysis the 4% reduction in performance produced by filling the postcue interval was not significant, F (1, 94) = 3.73, $p > .05$. The Cue × Filler interaction was not significant, F (3, 282) = 2.25, $p > .05$, but the means are shown in Table 1. When the postcue interval was filled, the 27% Only effect was significant, F (1, 282) = 40.80, $p < .01$. The ITBF recall did not differ from "either" recall, F (1, 282) < 1, and UTBF recall was significantly lower than ITBF recall, F (1, 282) = 30.42, $p < .01$. When the postcue interval was unfilled, the 33% Only effect was significant, F (1, 282) = 63.47, $p < .01$. The ITBF recall did not differ from "either" recall, F (1, 282) = .53, $p > .05$, and UTBF recall was significantly lower than ITBF recall, F (1, 282) = 61.98, $p < .01$.

Error data for 96 criterial Ss.—Fifteen % of the errors were omissions and 85% were intrusions. A large proportion (.91) of the intrusions were intralist intrusions. Table 2 shows the proportion of intralist intrusions which were members of the TBR and TBF sets for the filled and unfilled postcue interval. With an "either" cue, *S* was responsible for both sets and the labels TBR and TBF merely designate the tested and not-tested sets, respectively. When a TBF test was given, the label TBR refers to the set designated by the cue and the label TBF refers to the set actually

TABLE 2

Proportion of Intrusions Originating in To-Be-Remembered (TBR) Set and To-Be-Forgotten (TBF) Set

Cue cond.	Filled		Unfilled	
	TBR	TBF	TBR	TBF
Only	.73	.27	.81	.19
Either	.59	.41	.53	.47
ITBF	.27	.73	.22	.78
UTBF	.83	.17	.88	.12

Note.—For the "only" cue the TBR set is the cued set; for the "either" cue TBR is the tested set; for the ITBF cue TBR is the originally cued set; TBF is the tested (finally cued) set; for the UTBF cue TBR is the cued set, TBF is the actually tested set.

tested. For all four cues, there was a greater opportunity for intrusions from the set that was not tested because it contained two untested items; the set which was tested contained only one untested item. In order to correct for differential opportunity, mean numbers of intrusions were substituted for total numbers and proportions were then calculated. The proportions shown in Table 2 thus reflect the intrusion rate per item.

The error data provide strong evidence that *S* can maintain a selective search. When an "only" cue was given, a small proportion of the errors originated from the TBF set. When an "either" cue was given, however, almost half of the errors were TBF intrusions. The error data for the two types of TBF tests are especially impressive. On UTBF trials, the majority of the errors come from the set designated by the "only" cue. On ITBF trials, when the cue is countermanded by the asterisk of the test slide, the majority of intrusions stem from the set indicated by the asterisk.

Questionnaire data for the 96 criterial Ss.—Although all 96 *S*s appeared not to have changed their performance strategy on the experimental lists as determined by their verbal reports and the performance criterion, 50 (52%) of the *S*s claimed they found occasional tests of TBF items distracting. Since *S*s were not told that they would be given several UTBF tests, they were also asked if they found any of the cues to be misleading, and 47 (49%) of the *S*s claimed they had. Of these 47 *S*s, 32 correctly noted the nature of the deception; i.e., they stated that at least on one occasion the test item was not a member of the set designated by the cue and was not accompanied by an asterisk.

Most of the *S*s (87%) indicated that they tried to learn the pairs by remembering the associative link; the remainder simply tried to learn the responses. When asked if they could rehearse the pairs during the filled postcue interval, 12% replied yes, 43% replied no, and 45% replied that they could rehearse on some trials, especially if the subtraction problem was easy. When asked if they rehearsed during the unfilled postcue interval, 56% responded yes, 7% responded no, and 37% stated that they rehearsed during some of the unfilled intervals. When they were able to rehearse, 92% reported rehearsing the set(s) designated by the cue.

Discussion

The fundamental premise of the selective search hypothesis is that the cue exercises a directive influence on the search process. Two findings of the present experiment provide strong confirmation of this premise. First, there is the low level of UTBF recall relative to recall under the other cue conditions. When *S* received an immediate direction to the appropriate set ("only" cue), recall averaged 77%, when *S* received a delayed direction to the appropriate set (ITBF probe) recall averaged 44%, but when the cue misdirected *S*'s search (UTBF probe) recall was sharply down to 16%. It is clear that the low level of UTBF recall represents a decrement in performance, since in the absence of directions to search selectively ("either" cue) recall was at 48%. The error data complement the recall data. On the assumption that the locus of origin of the intrusions identifies the regions of search, we can claim evidence that *S* searches selectively in the set that has been specified by the cue.

The foregoing observations establish that cue-directed selective search occurs, but more than these observations is required to sustain the proposition that selective search is responsible for Only effect. Compelling evidence would have been present if ITBF recall had equaled "only" recall while "either" and UTBF had remained at their obtained levels. Instead, ITBF recall was significantly lower than "only" recall and equivalent to "either" recall. An appeal to selective rehearsal may seem to offer an account of the superior "only" recall, since on ITBF tests the introduction of the countermanding cue was concurrent with the probe item, leaving no time for selective rehearsal of TBR items. Although it is not possible to discard this interpretation, there is no evidence to support it. In particular, we would expect, on this account, that the difference between "only" and ITBF recall would be greater in the unfilled compared to the filled condition. The former condition provides greater opportunity for selective rehearsal of TBR items on "only" trials and selective rehearsal of the inappropriate set on ITBF trials. The data do not confirm this expectation. The Cue × Filler interaction was not significant and ITBF recall in the unfilled condition was 5% higher than in the filled condition.

The lower-than-expected ITBF recall could also be attributed to disruptive effects of the ITBF sequence. Assume that on the "only" cue *S* readies himself to search the designated set; the countermanding signal disrupts this state of readiness and requires *S* to redirect the search. On the further assumption that disruption has a general detrimental effect on performance, lower ITBF recall can be understood. As with the rehearsal interpretation there is no support for the disruption account in the experimental data. For example, the distribution of intrusions between the TBR and TBF sets is almost identical for the "only" and ITBF cue conditions (see Table 2). Thus, there is no evidence that on ITBF trials *S*s were less successful in maintaining a selective search.

Failure to confirm the prediction of equality of recall for "only" and ITBF cue conditions creates difficulties for the selective search hypothesis. Perhaps the aforementioned factors are responsible; however, in the absence of positive evidence in support of such an interpretation, we must admit the argument that mechanisms other than selective search may be involved in the Only effect. For example, a mechanism that is triggered by the original "only" cue and operates to reduce the availability of TBF items would cause lower TBF recall on ITBF trials relative to TBR recall on "only" trials. While it is clear that postulating an erasure mechanism would help explain the TBR-ITBF recall difference, it is also clear that other aspects of our data seem incompatible with an erasure mechanism. It is difficult to see how erasure can explain the finding that ITBF recall was consistently higher (ranging from 19% to 39% higher) than UTBF recall. Nor is it useful to propose that both selective search and erasure are involved since there seems no way of assessing the exact contributions of the two mechanisms. Given our earlier evidence (Epstein et al., 1972) and the evidence of almost perfect TBF availability in the experiments of Block (1971) and Elmes et al. (1970), we believe that further efforts to establish the full sufficiency of the selective search hypothesis are in order before more complicated, multifactor interpretations are considered.

A secondary aim of the experiment was to reexamine the influence of selective rehearsal. Our earlier studies (Epstein et al., 1972; Shebilske et al., 1971) showed that the magnitude of the Only effect was independent of our rehearsal manipulations. The within-*S* design of the present experiment was intended to provide better conditions for detecting a rehearsal effect. The analysis of the original 96 *S*s, which included noncriterial *S*s, revealed a significantly greater Only effect for the unfilled condition both on the training and experimental trials. The difference was entirely due to superior "only" recall on unfilled trials relative to filled trials. Since opportunity for rehearsal in the unfilled trials would be expected to favor selective rehearsal of TBR items on "only" trials, this outcome is entirely consistent with the operation of selective rehearsal. A significant Filler × Cue interaction was also obtained on the training trials for the final 96 criterial *S*s. Again the Only effect was larger in the unfilled condition and again this was attributable to higher "only" recall in the unfilled condition. Consistent with this latter finding were *S*s' reports in the postexperimental inquiry that when they did rehearse they devoted themselves almost exclusively to the cued set.

In our earlier studies (Epstein et al., 1972; Shebilske et al., 1971) we concluded that although selective rehearsal is not a necessary condition of the Only effect, there is no reason to deny that rehearsal can enhance the effectiveness of the cues. There are a number of ways in which this could come about. One possibility that would fit in with our emphasis on selective search is that maintenance of the cued set in active memory through rehearsal insures that the cued set will be the search set when the probe stimulus is presented. Selective rehearsal is an implied decision to search selectively in the rehearsed set. Accordingly, the Only effect will be greater when selective rehearsal occurs.

It has become common to distinguish between the structural components of memory and the control processes of memory. The latter include such processes as rehearsal and retrieval strategy and are considered to be under *S*'s voluntary control. The directed forgetting experiment assumes that control processes figure prominently in recall. The evidence of selective search provided by the present experiment fully justifies that assumption.

REFERENCES

Archer, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, **74** (10, Whole No. 497).

Battig, W., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 1969, **80** (3, Pt. 2).

Bjork, R. A. Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 255–268.

Block, R. A. The effects of instructions to forget in short-term memory. *Journal of Experimental Psychology*, 1971, **89**, 1–9.

Elmes, D. G., Adams, C., & Roediger, H. Cued forgetting in short-term memory: Response selection. *Journal of Experimental Psychology*, 1970, **86**, 103–107.

Epstein, W. Poststimulus output specification and differential retrieval from short-term memory. *Journal of Experimental Psychology*, 1969, **82**, 168–174.

Epstein, W. Facilitation of retrieval resulting from postinput inclusion of part of the input. *Journal of Experimental Psychology*, 1970, **86**, 190–195.

Epstein, W. Mechanisms of directed forgetting. In G. Bower (Ed.), *The psychology of learning and motivation*. Vol. 6. New York: Academic Press, in press.

Epstein, W., Massaro, D. W., & Wilder, L. Selective search in directed forgetting. *Journal of Experimental Psychology*, 1972, **94**, 18–24.

Harley, W. F. The effect of monetary incentive in paired associate learning using an absolute method. *Psychonomic Science*, 1965, **3**, 141–142. (a)

Harley, W. F. The effect of monetary incentive in paired associate learning using a differential method. *Psychonomic Science*, 1965, **2**, 377–378. (b)

Harley, W. F. Delay of incentive cues in paired associate learning and its effect on organizing responses. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 924–929.

Loftus, G. R., & Wickens, T. D. Effect of incentive on storage and retrieval processes. *Journal of Experimental Psychology*, 1970, **85**, 141–147.

Montague, W. E., Kiess, H. D., Hillix, W. A., & Harris, R. Variations in report of covert rehearsal and in STM produced by differential payoff. *Journal of Experimental Psychology*, 1970, **83**, 249–254.

Reitman, W. R., Malin, J. T., Tanner, J. E., Bjork, R. A., & Higman, B. *Strategy control in directed forgetting*. (Mental Health Research Institute Communication No. 284) Ann Arbor: University of Michigan, 1971.

Shebilske, W., Wilder, L., & Epstein, W. Forget instructions: The effect of selective rehearsal and categorical distinctiveness. *Journal of Experimental Psychology*, 1971, **89**, 372–378.

Wickens, D. D., & Simpson, C. K. Trace cue position, motivation and short-term memory. *Journal of Experimental Psychology*, 1968, **76**, 282–285.

(Received January 18, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 358–364

THE PONZO ILLUSION IN STEREOSCOPIC SPACE[1]

R. T. GREENE, R. B. LAWSON, AND CYNTHIA L. GODEK[2]

University of Vermont

Eight *O*s reported stereoscopic size and distance of the enclosed lines of a Ponzo configuration as a function of their magnitude and direction of disparity. The magnitude of the illusion varied directly with the apparent size of the disparate lines, and the illusion persisted even when the critical contours appeared in different stereoscopic depth planes.

Recently, there has been renewed interest in the Ponzo illusion because it is claimed to be prototypic of geometrical-optical illusions which employ the perspective cue to the perception of distance, and more importantly because the illusion is said to arise from the application of the normal size constancy mechanism in an inappropriate situation (Gregory, 1966; Leibowitz & Judisch, 1967a; Tausch, 1954).

Figure 1 presents the geometrical or classical Ponzo configuration, and it is apparent that the apical vertical line appears longer despite the fact that both vertical lines are of equal physical extent and, therefore, produce identical retinal images. Generally, explanations of the Ponzo illusion suggest that the converging lines or wedge falsely signal distance and thereby induce a misapplication of the mechanism underlying size constancy (Gregory, 1966; Tausch, 1954). Essentially, the "misapplication of size constancy" hypothesis, first suggested by Thiery (1896) to explain the Muller-Lyer illusion, states that an hypothetical distance cue (arising from linear perspective) gives rise to a perceptual distortion in the size of the apical vertical line. Presumably, this distortion occurs because if the apical line (constant angular size) actually did lie at a greater distance from *O* than the other physically equal vertical line, the former would have to be physically larger to maintain its fixed angular size. According to Gregory, the *apparent* distance changes consistent with the "misapplication of size constancy" hypothesis are prohibited by the restriction of the Ponzo configuration to a two-dimensional surface.

Unfortunately, despite its impact, systematic empirical evidence supporting the perspective explanation of the Ponzo illusion is marginal. For example, the magnitude of the Ponzo illusion and size constancy for distant objects increase from childhood through adulthood, whereas the magnitude of the Ponzo illusion decreases while the tendency toward size constancy remains the same in old age (Leibowitz & Judisch, 1967a, 1967b; Zeigler & Leibowitz, 1957). Moreover, Leibowitz, Brislin, Perlmutter, and Hennessy (1969) hypothesized that if the Ponzo wedge serves as a cue to distance, then differential experience with perspective depth cues should influence the magnitude of the Ponzo illusion. They obtained size judgments consistent with the above argument provided that the Ponzo configuration contained additional cues to distance (e.g., texture gradients); however, they found no significant difference in the magnitude of the Ponzo illusion between their American and Guamanian *S*'s for the classical stimulus configuration (Fig. 1). In fact, the Guamanian *S*s, who supposedly had fewer experiences with perspective cues, showed a slighlty larger illusion for the Fig. 1 configuration.[3]

Recently, Fisher (1968a, 1968b; 1969)

---

[1] This research was supported in part by Grants USAFOSR-F 44620-69-C-0004 and NSF GB 30579 to R. B. Lawson, and NASA Traineeship NS6T-28-52 to R. T. Greene.

[2] Requests for reprints should be sent to R. T. Greene, Department of Psychology, John Dewey Hall, University of Vermont, Burlington, Vermont 05401.

[3] Contrary to the perspective theory of optical illusions, Over (1968) has shown tactile counterparts of some of the geometrical-optical illusions, even though no perspective cues (due to projections) exist in tactile space.

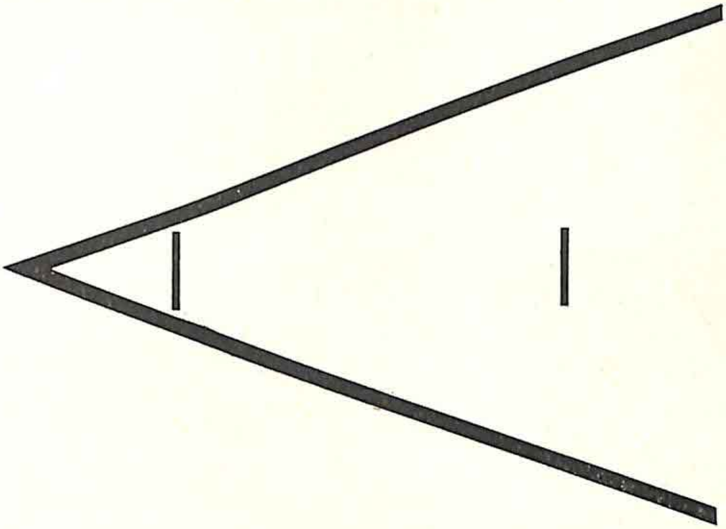


FIG. 1. The classic or geometric Ponzo configuration. (Even though both lines produce identical retinal images, the apical line appears larger.)

has proposed that it is the *mere* presence of the converging lines themselves rather than a hypothetical distance cue provided by the wedge which gives rise to the illusion. Fisher (1968b) has shown that the magnitude of the Ponzo illusion varies inversely with the convergence angle of the wedge. Also, he found that inserting more than two vertical lines within the wedge results in a gradient of apparent size distortion of the enclosed lines, and that the distortion gradient arises even when only one of the converging lines is presented.[4] Finally, Fisher (1968a) has demonstrated that the illusion persists when the Ponzo wedge is replaced by pairs of enclosing *parallel* lines of differential proximity to the enclosed vertical lines. Accordingly, the magnitude of the size distortion in the Ponzo illusion apparently depends upon the proximity of the contours defining the wedge and the vertical lines. The results from line-frame articulation studies (Kunnapäs, 1955), which indicate that the area of a frame influences markedly the apparent length of an enclosed line, are consistent with a contour adjacency explanation of the Ponzo illusion.

With few exceptions (Klix, 1962; Thiery, 1896), geometrical-optical illusory configurations have been presented upon two dimensional surfaces, and as a consequence experimental manipulations of the contours of the configuration have been restricted to flat planar displays. Previous investigations employing stereoscopic presentation of the Ponzo configuration (wedge to one eye, vertical lines to the other eye) indicate that the illusion persists even though the stereoscopic pattern is labile due to the absence of corresponding contours in the composite stereoscopic field (Fisher & Lucas, 1970). In fact, any apparent separation of stereoscopic depth between the wedge and the enclosed vertical lines was minimized because such a separation purportedly reduced the illusory effect (Day, 1961; Schiller & Wiener, 1962).

Contrariwise, in the present experiment, the complete Ponzo configuration was presented in both half-images so that the magnitude and direction (in front of or behind the wedge) of the enclosed vertical lines could be varied as a function of retinal disparity. Whereas previous investigators employed dichoptic techniques to determine the neural locus of origin of illusions, we determined if the Ponzo illusion would vary as a function of the magnitude and

---

[4] In fact, W. Sickles (1942) was first to demonstrate the apparent size distortion gradient which results from differential locations of the vertical lines within the Ponzo wedge.

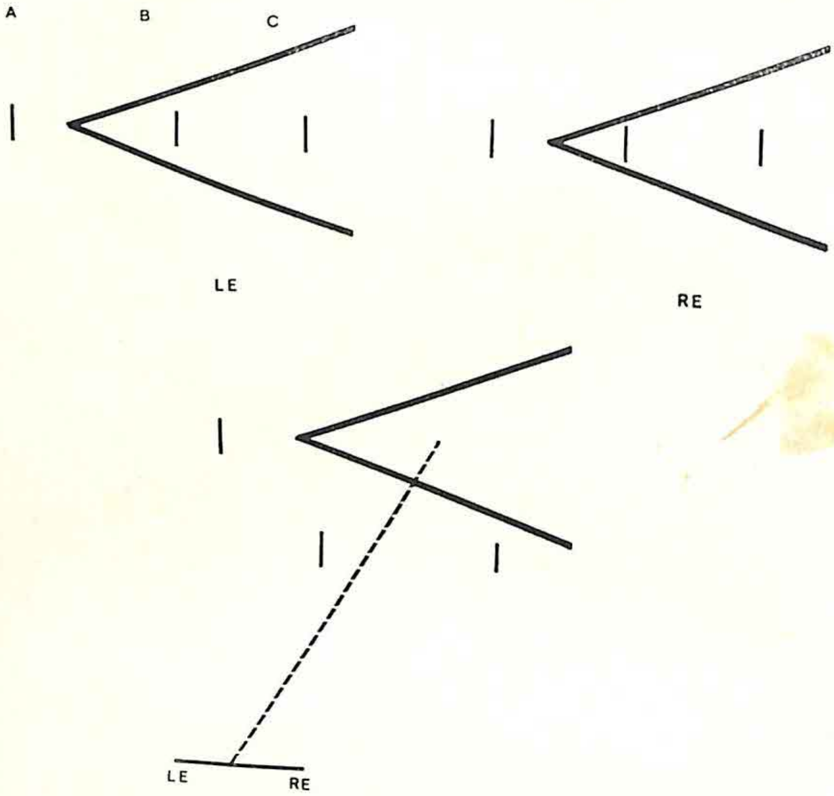


FIG. 2. Experimental stereogram and a schematic representation of the perceptual results produced by viewing stereoscopically the half-images. (In this example, the enclosed vertical lines carry an equivalent positive (crossed) disparity, and consequently they appear in front of the wedge.)

direction of disparity carried by the enclosed vertical lines, and, therefore, if contour adjacency influences the magnitude of the illusion only when the entire Ponzo configuration is confined to two dimensions. This procedure represents a methodological departure from previous stereoscopic investigations of illusions inasmuch as we manipulated equally the apparent distance and size of both enclosed contours as a function of disparity in order to vary their apparent adjacency to the enclosing contours along the y (height) and z (distance) axes of stereoscopic space.

METHOD

Observers.—Eight females (mean age = 22.7 yr.) served as *O*s. Each had uncorrected normal near visual acuity socres (20/17–20/25), normal lateral (3–6) and vertical (6–9) phoria, and stereopsis (90% Fey-Shepard) as measured by a Bausch & Lomb Orthorater (Model 71-21-31). Although each had experience with stereoscopic phenomena, none was familiar with the purpose of the experiment.

Apparatus.—Three types of stereograms were employed. One type, used to train *O*s in making size and distance magnitude judgments, consisted of five stereograms containing only three vertical lines (angular size constant at 41.5′), Lines A, B, and C, in which either Line B *or* C carried a crossed or uncrossed disparity.

Figure 2 presents a prototype of the five experimental stereograms employed, and a schematic representation of the stereoscopic organization of the half-images. Both of the enclosed vertical lines in each of the experimental stereograms carried *equivalent* disparities, and the disparity varied from +38.3′ (crossed) thru zero to −38.3′ (uncrossed) in steps of 19.15′ of arc at a viewing distance of 81 cm. The disparities were produced so that both vertical lines always carried the same visual direction and magnitude relative to the wedge when the half-images were in binocular register. In other words, the disparities of the enclosed lines in each half-image were generated by lateral displacement of these lines in opposite directions from a central point. Thus, differences in the stereoscopic size of the enclosed lines as a function of the direction of disparity were not confounded by the contour adjacency of the enclosed lines and the wedge within the planar half-images.

Lastly, the five control stereograms were identical to the experimental patterns except that they consisted of *only* the three vertical lines. The angular sizes of all three vertical lines in all of the stereograms were equivalent and remained constant (41.5′). Finally, both types of stereograms were displayed in a prism stereoscope.

Procedure.—All *O*s first estimated depth and size in the training stereograms. Under stereoscopic conditions, each *O* was instructed to consider the distance between themselves and Line A, which was always nondisparate, as equal to 100 units, and then to estimate the distance of Lines B and C relative to Line A. To estimate size, Line A served as a referent of 10 units against which the lengths of B and C were estimated. Random presentation of the five training stereograms yielded reliable stereoscopic size and distance judgments from each *O*.

Both the five experimental and five control stereograms were presented four times to each *O* as randomized blocks which were counterbalanced across *O*s. Each *O* was instructed to estimate the size and distance of Lines B and C in the same fashion as practiced previously.

RESULTS

The results indicate clearly that the Ponzo illusion occurred even when the wedge and the enclosed vertical lines appeared in different stereoscopic depth planes. Figure 3 presents the mean size estimates for Lines B and C in the presence and absence of the wedge based upon all eight *O*s across the four trials. Note that with increasing positive disparity the vertical lines appeared progressively smaller to *O*s, whereas they grew progressively larger with increasing negative disparity. An analysis of variance of the mean size estimates was computed, and indicated that disparity, F (4, 28) = 39.32, $p < .01$; line position within the wedge, F (1, 7) = 22.69,

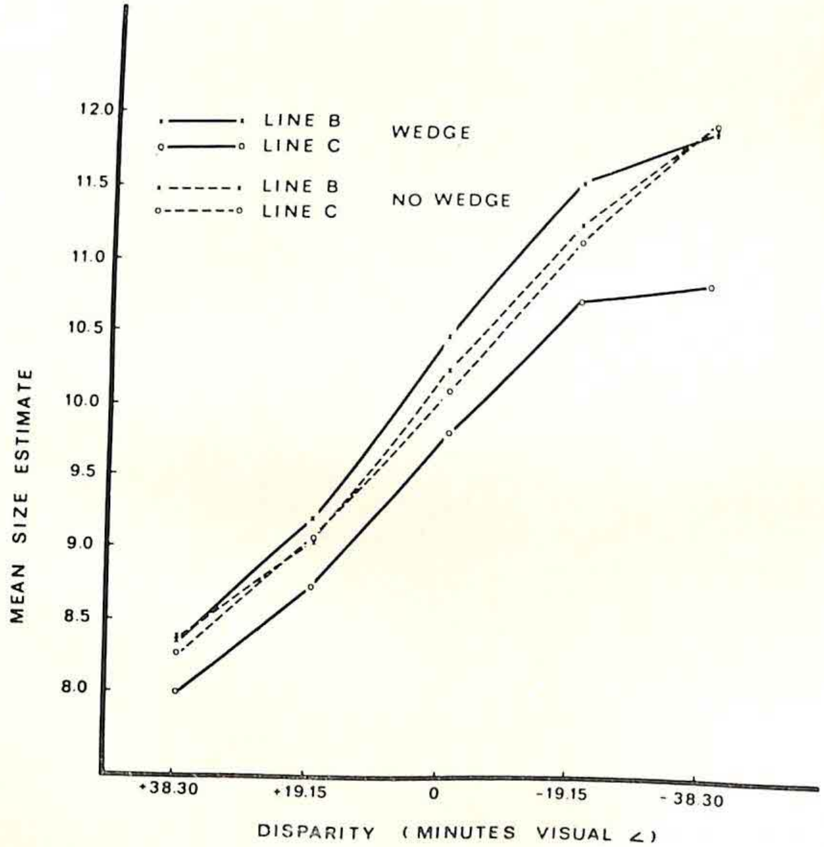


FIG. 3. The mean size estimates based upon all eight *O*s across the four trials as a function of the magnitude and direction of disparity with and without the Ponzo wedge in the stereoscopic field.

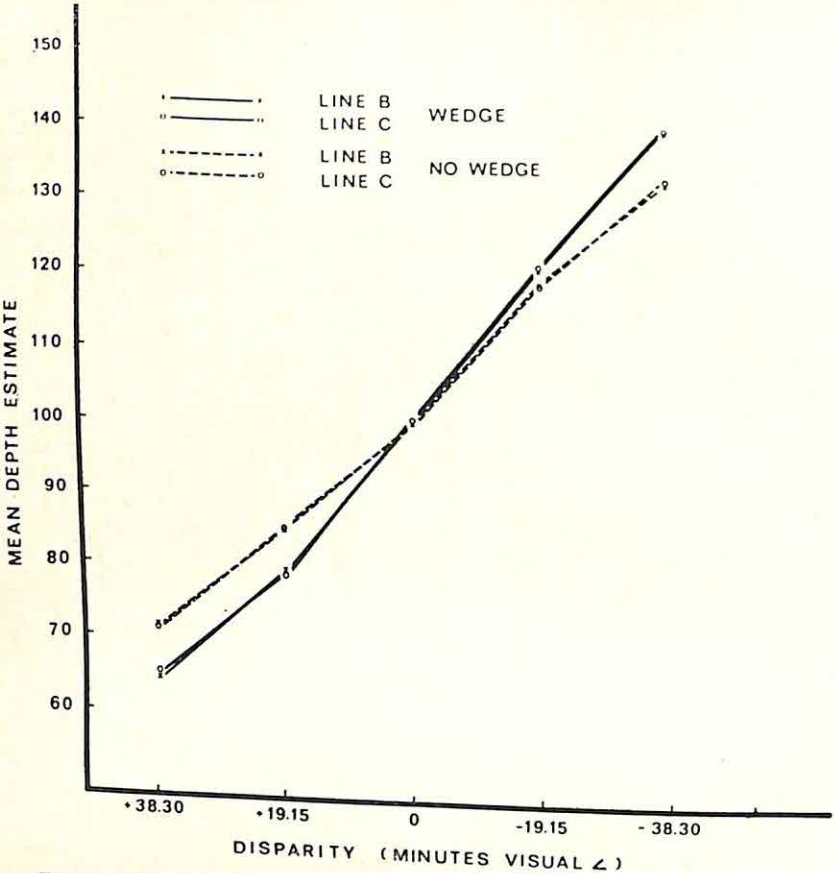


FIG. 4. The mean depth estimates based upon all eight *O*s across the four trials as a function of the magnitude and direction of disparity with and without the Ponzo wedge in the stereoscopic field.

$p < .01$; Line Position × Wedge interaction, F (1, 7) = 21.53, $p < .01$; and the three-way Disparity × Line Position × Wedge interaction, F (4, 28) = 5.33, $p < .01$ affected significantly the stereoscopic size of Lines B and C. Thus, the perceived parity; however, the magnitude of the illusion increased as the lines appeared larger in the stereoscopic field.

Figure 4 presents the mean distance estimates for Lines B and C for the experimental and control stereograms based upon all eight *O*s across the four trials. A depth estimate of less than 100 units indicates that the enclosed lines appeared in front of the wedge, whereas, depth estimates greater than 100 units indicate that they appeared behind the wedge. The only significant treatments affecting the stereoscopic distance of the enclosed lines were disparity, F (4, 28) = 14.55, $p < .01$; and the Wedge × Disparity interaction, F (4, 28) = 3.33, $p < .05$. Thus, the stereoscopic distance of the enclosed lines varied directly with disparity, and the wedge appeared to facilitate the judgment of such distances.

Finally, Fig. 5 presents the percent Ponzo illusion based upon all eight *O*s as a function of disparity carried by the enclosed vertical lines.[5] Recall that as both vertical lines appeared progressively closer to *O* they also became progressively shorter, whereas they grew progressively longer with their increasing distance from *O*. Thus, when the enclosed vertical lines appeared in different stereoscopic depth planes from the wedge as a function of disparity, the changes in *apparent* line length altered the phenomenal adjacency (y axis) between the lines and the wedge. The fact that the wedge alone did not affect significantly size estimates, $F\ (1, 7) = 3.68, p > .05$, but rather interacted with line position and disparity to alter significantly stereoscopic size estimates of the enclosed lines supports this conclusion.

---

[5] The percent illusion equals the difference in mean size estimates between Lines B and C divided by the mean size estimate for Line B, times 100.

DISCUSSION

The present results demonstrate clearly that the Ponzo illusion arises in stereoscopic space, and that it persists even when the enclosed and enclosing contours appear in different depth planes. Furthermore, the present findings indicate that the magnitude of the Ponzo illusion varies directly with the magnitude *and* direction of disparity carried by the enclosed vertical lines.

As a result of increasing positive disparity, the enclosed vertical lines appeared progressively closer and smaller to *O*s, and, consequently, the apparent fronto parallel adjacency between the vertical lines and the wedge decreased (y axis). Likewise, the depth adjacency between these contours along the z axis of stereoscopic space was reduced (Gogel & Mershon, 1969). Contrariwise, the enclosed lines appeared progressively larger and farther to *O*s with increasing negative disparity, and, therefore, the apparent contour adjacency between the enclosed vertical lines and the wedge *increased* although the depth adjacency decreased. Thus, the fact that the magnitude of the illusion varied directly with the perceived size of the vertical lines agrees

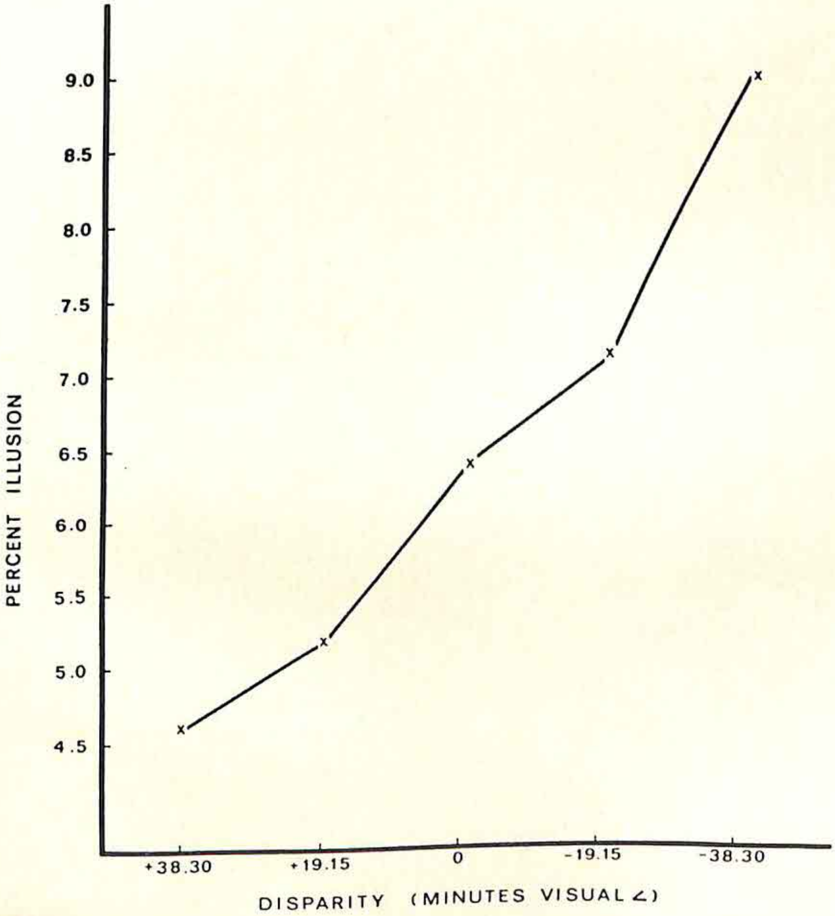


FIG. 5. The percent of the Ponzo illusion based upon all eight *O*s across the four trials as a function of the magnitude and direction of disparity carried by the enclosed vertical lines.

with a contour adjacency explanation of the Ponzo illusion even when the critical contours appear in different stereoscopic depth planes.

Recently, Gogel and Mershon (1969) reported that the magnitude of the Gelb whiteness contrast effect decreased with an increase of stereoscopic depth between their test and inducing disks. Unfortunately, their targets carried only crossed or positive disparities and had we employed only positive disparities we would have concluded similarly that a reduced Ponzo illusion results when the enclosed and enclosing contours appear in different depth planes. However, our data clearly indicate that uncrossed or negative disparities yielded a larger Ponzo illusion than the zero or positive disparity conditions.

Finally, Julesz (1971) has portrayed the Ponzo configuration by random-dot stereograms, and reports that the illusion is reduced when the wedge and enclosed lines appear in different stereoscopic depth planes. Unfortunately, all of his targets contained only positive or crossed disparities, and although in one instance the wedge and enclosed lines appeared in different stereoscopic depth planes no systematic data was presented to support his observations of a reduced illusory effect.

In summary, the present results agree with previous findings that the Ponzo illusion persists even when the enclosing and enclosed contours appear in different stereoscopic depth planes. However, the present findings indicate that the magnitude of the illusion varies directly with the magnitude and direction of the disparity carried by the enclosed lines. Furthermore, our findings suggest that the contour adjacency explanation of the illusion can accommodate the results obtained from both two- and three-dimensional Ponzo configurations regardless of whether changes of contour adjacency are real or apparent.

REFERENCES

Day, R. H. On the stereoscopic observation of geometrical illusions. *Perceptual and Motor Skills*, 1961, **13**, 247–258.

Fisher, G. H. An experimental comparison of rectilinear and curvilinear illusions. *British Journal of Pshchology*, 1968, **59**, 23–28. (a)

Fisher, G. H. Gradients of distortion seen in the context of the Ponzo illusion and other contours. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 212–217. (b)

Fisher, G. H. Towards a new explanation for the geometrical illusions: I. The properties of contours which induce illusory distortion. *British Journal of Psychology*, 1969, **60**, 179–185.

Fisher, G., & Lucas, A. Geometrical illusions and figural after-effects. The distorting and distorted components of illusions. *Vision Research*, 1970, **10**, 393–404.

Gogel, W., & Mershon, D. H. Depth adjacency in simultaneous contrast. *Perception & Psychophysics*, 1969, **5**, 13–17.

Gregory, R. L. *Eye and brain*. New York: McGraw-Hill, 1966.

Julesz, B. *Foundation of cyclopean perception*. Chicago: University of Chicago Press, 1971.

Klix, F. *Elementaranalysen zur psychophysik der Raumwahrnehmung*. Berlin: Deutscher Verlag der Wissenschaften, VEB, 1962.

Kunnapäs, T. M. Influence of frame size on apparent length of a line. *Journal of Experimental Psychology*, 1955, **50**, 168–170.

Leibowitz, H., Brislin, R., Perlmutter, L., & Hennessy, R. Ponzo perspective illusion as a manifestation of space perception. *Science*, 1969, **166**, 1174–1176.

Leibowitz, H. N., & Judisch, J. M. The relation between age and the magnitude of the Ponzo illusion. *American Journal of Psychology*, 1967, **80**, 105–109. (a)

Leibowitz, H. W., & Judisch, J. M. Size-constancy in older persons: A function of distance. *American Journal of Psychology*, 1967, **80**, 294–296. (b)

Over, R. Explanations of geometrical illusions. *Psychological Bulletin*. 1968, **70**, 545–562.

Schiller, P., & Wiener, M. Binocular and stereoscopic viewing of geometric illusions. *Perceptual and Motor Skills*, 1962, **15**, 739–747.

Sickles, W. Experimental evidence for the electrical character of visual fields derived from a quantitative analysis of the Ponzo illusion. *Journal of Experimental Psychology*, 1942, **51**, 84–89.

Tausch, R. Optische Täuschungen als artifizielle Effekte der Gestaltungsprozesse von Grössen— und Formenkonstanz in der naturlichen Raumwahrnemung. *Psychologie Forschung*, 1954, **24**, 299–348.

Thiery, A. Über geometrisch—optische Tauschungen. *Philosophische Studien*, 1896, **12**, 67–125.

Zeigler, H. P., & Leibowitz, H. Apparent visual size as a function of distance for children and adults. *American Journal of Psychology*, 1957, **70**, 106–109.

(Received January 26, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 365–369

EFFECTS OF "RIGHT" AND "WRONG" IN A THORNDIKIAN EXPERIMENT[1]

IN-MAO LIU[2]

National Taiwan University, Taiwan, Republic of China

The effect of RIGHT and WRONG in a modified Thorndikian experiment was compared with that of the initially neutral outcomes of ARMY and NAVY. Under the latter condition, *S* was told that ARMY signaled RIGHT or WRONG and NAVY signaled WRONG or RIGHT, but not which is which. After these information trials were all over, *S* was told which represented which. There was a third condition of giving *S* information about his response, in which the outcomes of ARMY and NAVY were used and *S* was told which (ARMY or NAVY) represented which (RIGHT or WRONG) before the information trials. More repetitions of correct responses were found only in the RIGHT–WRONG condition, and with training the repetitions of correct responses increased also significantly. It was concluded that A. M. Buchwald's 1969 formulation explains only part of the results of a Thorndikian experiment.

There have been a number of experiments attempting to assess the effect of *E*'s announcing RIGHT or WRONG following *S*'s responses in the Thorndikian type of multiple-choice verbal learning experiments (reviewed in Postman, 1962). Recently, Buchwald (1969) has advanced a view that on a test trial in a Thorndikian experiment RIGHT responses are repeated more often than WRONG responses because for some stimuli *S* recalls both his previous responses and their outcomes, and he repeats the responses when the outcome was RIGHT and does not when it was WRONG. Let us consider one implication of this view.

Suppose that at the start of an experiment *S* is informed as to the response alternatives available to him, is told that he must make one of these responses each time a stimulus item is presented, and is instructed to try to learn which response is correct for each item. We assume that two trials are used, each trial consisting of one presentation of every stimulus item. The first trial is an information trial on which *E* either says ARMY or NAVY, following each response. The *S* has previously been informed that one of these (ARMY or NAVY) represents RIGHT and that the other represents WRONG, but not which is which. Before the start of the second trial, which is a test trial, *S* is told which of these (ARMY or NAVY) represents RIGHT and which represents WRONG. On this test trial, *S* is instructed to try to give correct responses but is still required to make some response to each item as it is presented.

According to the view that RIGHT responses are repeated more often than WRONG responses because for some stimuli *S* recalls both his previous responses and their outcomes, and he repeats the responses when the outcome was RIGHT and does not when it was WRONG, the above hypothetical experiment should yield results similar to the usual experiment in which straightforward RIGHT and WRONG are announced on the information trial. This is because the probability of recalling ARMY or NAVY following *S*'s response in the hypothetical experiment would be about the same as the probability of recalling RIGHT or WRONG in the usual experiment, although this is an additional assumption which is not contained in Buchwald's (1969) formulation.

This study reports such a hypothetical experiment with slight modifications. First, two information trials were administered prior to the test trial. In this way it was possible not only to study the probability of repeating the previous response to a given stimulus, but to study a change in this probability as a consequence of training. Second, another condition was introduced, in which *E* also announced ARMY or NAVY following *S*'s response, but *S* was told in advance which of these represented RIGHT and which represented WRONG. This condition was introduced on the following grounds. If the RIGHT–WRONG condition produces effects different from those of the ARMY–NAVY condition, the former effects may have been acquired through *S*'s past experience. Then, by introducing the ARMY–NAVY condition in the laboratory, *S* can acquire the new roles attached to ARMY and NAVY as a child first learns the meanings of RIGHT and WRONG. In this way, it may be possible to study the genesis or process of the effects of rewards and punishments in human learning. Third, in a typical Thorndikian experiment, *S* responds to a stimulus word by selecting one out of several response alternatives. In the present experiment, the response alternatives were A and C. In this case, *S* might choose A all the time, or he might choose A or C, depending on the spelling pattern of a stimulus word, etc. This may make a Thorndikian experiment unnecessarily complicated. Therefore, in the following experiment, on the first information trial *S* was to respond to a sequence of stimulus words with a prearranged response sequence.

---

[1] This study was supported by a grant from the National Science Council. The assistance of Hsen-hsing Ma, Sin-tsi Lin, Chung-ing Jeng, and Shui-ju Yang is gratefully acknowledged. The author is also grateful to Charles N. Cofer, Ovid Tzeng, and Ellen Yang for their reading and critical comments.

[2] Requests for reprints should be sent to In-Mao Liu, Department of Psychology, National Taiwan University, Taipei, Taiwan, Republic of China.

METHOD

Subjects and materials.—The *S*s were 180 students from introductory courses in psychology. Forty words were randomly selected from the first 150 words in the Liu-Chuang (1970) list. Since the Chinese words in this list were arranged in the order of meaningfulness value, they were all highly meaningful words.

The words were randomly assigned to one of two 20-word lists. As in the experiments of Buchwald (1962), for each list, 4 words were held fixed in Positions 1, 2, 19, and 20, and the other words were arranged in three different random permutations. The different permutations were used in different trials with the same list.

Procedure.—The *E* read a set of instructions at the start of the experiment. Portions of these instructions were reviewed before each trial. The *S* was told that he would be read a list of words, that one of the letters A or C had been randomly assigned to each word, that his task was to learn which letter went with which word. On the first information trial, *S* was instructed to say one of the two letters, each time *E* said a word, according to a prearranged order explained below.

There were three conditions for giving *S* information about his response. Under one condition (Cond. RW), *S* was told that *E* would indicate a correct (incorrect) response by saying RIGHT (WRONG). Under another condition (Cond. AN), *E* said ARMY or NAVY each time *S* responded. Only after the information trials were all over, did *E* tell *S* that ARMY (or NAVY) signaled *S*'s response being correct and NAVY (or ARMY) signaled *S*'s response being incorrect. Which response outcome signaled correctness was randomized. Under the third condition (Cond. AN-RW), the procedure was the same as in Cond. AN, but *S* was told in advance which of ARMY or NAVY signaled RIGHT and which signaled WRONG.

Actually there were no correct responses. The words were randomly divided into two equally large sets. On the RIGHT words, *E* said RIGHT (or ARMY or NAVY) to every response. On the WRONG words, *E* said WRONG (or NAVY or ARMY) to every response. Equal numbers of RIGHT words and WRONG words appeared in the first 2, last 2, and middle 16 positions.

The *S*s were divided into six groups of 30 *S*s each according to order of their appearance to form a 3 × 2 factorial design. The first factor refers to the 3 conditions mentioned above; the second, to the number of information trials, which was either 1 or 2. For half of the *S*s of each group, one list was used; for the other half, another list was used. On the first information trial, each *S* was to say one of the letters (A or C) according to a prearranged order each time *E* said a word. This was accomplished by preparing a pile of 20 cards. On the back of each card was printed A or C. There were 10 A cards and 10 C cards arranged in a random order. For every 6 *S*s assigned to the six groups, one order was used. Therefore, there were 30 random orders. Each time *E* said a word, *S* turned a card and read the letter according to the order of cards.

For *S*s given only one information trial (Groups 1RW, 1AN, and 1AN-RW; the numeral indicating the number of information trials), a test trial then followed, on which *E* said nothing after *S*'s response. The third trial was a recall-of-outcome trial, on which, each time *E* said a word, *S* responded with RIGHT or WRONG or with ARMY or NAVY, depending on the group to which he belonged.

For *S*s given two information trials (Groups 2RW, 2AN, and 2AN-RW), they were to select their own responses on the second information trial, and *E* responded RIGHT or WRONG (or ARMY or NAVY) consistently from the first (1_1) to second (1_2) information trials. If a RIGHT response (ARMY or NAVY response for Groups 2AN and 2AN-RW) on Trial 1_1 was repeated on Trial 1_2, RIGHT (ARMY or

NAVY) was announced; if S responded otherwise, WRONG (NAVY or ARMY) was announced. If a WRONG response (NAVY or ARMY response) was repeated on Trial 1_2, WRONG (NAVY or ARMY) was announced; if S responded otherwise, RIGHT (ARMY or NAVY) was announced. A test trial then followed. On the test trial, a response was scored as a correct response or an incorrect response always in reference to Trial 1_1. For example, if a RIGHT response on Trial 1_1 was repeated on the test trial, it was scored as a RIGHT response. No recall-of-outcome trial was administered to any S given two information trials.

On the information and test trials, the rate of stimulus presentation was not controlled explicitly, but E attemped to minimize the time between S's response and the next succeeding stimulus presentation. The E recorded the time taken for each trial with a stopwatch. No trial took more than 90 sec.

RESULTS

Response repetitions.—The data of response repetitions consisted of the number of instances in which S gave the same response to a given stimulus word on Trial 1_1 and test trials. The symbols RR and WW will be used for response repetitions following RIGHT and WRONG or their equivalents, respectively. Table 1 shows the mean number of such repetitions expressed as proportions of the number of possible repetitions. The responses to the first two and last two words of a list were excluded from consideration.

From examination of Table 1 it can be seen that the effect of training was to increase RR repetitions. This effect was most pronounced under the RW condition. The WW data seem not to have been affected by training. With respect to the outcome conditions, the performance was most superior under the RW condition, with the performance level under the AN–RW condition falling in between those of the RW and AN conditions, when RR repetitions were used as a performance index. The RR repetitions under the AN condition were slightly less than those expected by chance. Perhaps, the prearranged orders of A or C responses did not favor RR occurrences. If this happened, it should have affected the performance under the other two conditions in the same way.

TABLE 1

MEAN PROPORTIONS OF RESPONSES REPEATED ON TRIAL 1_1 AND TEST TRIAL TO THE SAME STIMULUS FOLLOWING RIGHT, WRONG OR THEIR EQUIVALENTS

Information trial	Outcome cond.					
	RW		AN		AN-RW	
	RR	WW	RR	WW	RR	WW
1	.59	.43	.43	.49	.52	.48
2	.72	.48	.48	.48	.58	.42

An analysis of variance applied to the RR data showed that the training was a significant source of variance, F (1, 174) = 9.44, 0 < .01. The outcome conditions also differentially affected the RR repetitions, F (2, 174) = 20.84, $p < .001$. The interaction of training and outcome conditions was not significant, F (2, 174) < 1.

As to the three groups with a single information trial, more RR repetitions were obtained only in Group 1RW than in Group 1AN, t (174) = 3.63, $p < .001$. All the other differences were not significant, t (174) = 1.71 for Group 1RW versus Group 1AN-RW, and t (174) = 1.92 for Group 1AN-RW versus Group 1AN.

An analysis of variance applied to the WW data showed that neither the effects of training nor of outcome conditions were significant, F (1, 174) < 1, and F (2, 174) < 1, respectively. The interaction of training and outcome conditions was also not significant, F (2, 174) = 1.33.

Recall of outcomes.—Table 2 shows the proportion of outcomes correctly recalled for each group. According to this table, recall of RIGHT outcome or its indicator (ARMY or NAVY) was best in Group 1RW, next in Group 1AN-RW, and near a chance level in Group 1AN. An analysis of variance showed that the overall difference in recall of RIGHT outcome or its indicator was significant among these three groups, F (2, 87) = 3.35, $p < .05$. Further analyses showed that only the difference between Group 1RW and Group 1AN was significant, t (87) = 2.54, $p < .05$.

As is apparent from Table 2, recall of WRONG outcome or its indicator was near a chance level for all the three groups. An analysis of variance showed that the

TABLE 2

Mean Proportions of Correct Recall of Outcomes (Right, Wrong, or Their Equivalents)

Outcome cond.					
RW		AN		AN-RW	
R	W	R	W	R	W
.62	.48	.52	.53	.59	.50

overall difference in recall of WRONG outcome or its indicator was not significant among the groups, F (2, 87) < 1.

Discussion

There was less repetition following WRONG in the present experiment than is usually reported. One possibility is as follows. In the present experiment, since S had to respond to a series of stimulus words with a prearranged order of A and C responses, S's predisposition to respond with one of these (A or C) to a stimulus word could be controlled. If this type of S predisposition was responsible for a slightly higher level of repeating WRONG responses in a typical Thorndikian experiment, the WRONG responses repeated should be near a chance level irrespective of the outcome conditions in the present experiment. The present results are in line with this interpretation.

Since more correct responses were repeated under the RW condition than under the AN condition, Buchwald's (1969) model will have some difficulty in accounting for this finding. It may be argued that the probability of recalling ARMY or NAVY is lower than the probability of recalling RIGHT or WRONG following S's response. However, it has to be noted that according to the present results only the probability of recalling RIGHT was higher than the probability of recalling ARMY or NAVY (when it indicated RIGHT). The latter probability was near a chance level. As to the probabilities of recalling WRONG and its indicator, no difference was obtained. Therefore, an explanation based on differences in recall of RIGHT and WRONG on the one hand and ARMY and NAVY on the other can be ruled out.

The above consideration naturally leads to the following speculation. Let us illustrate this by means of a sequence of three events: the presentation of the stimulus, the making of the response, and the occurrence of the outcome. Imagine a sequence consisting of the stimulus SCHOOL, the response A, and the outcome RIGHT. Immediately following E's announcement of RIGHT, we conjecture that S rehearses the SCHOOL-A-RIGHT sequence. Although E attempted to minimize the time between S's response and the next succeeding stimulus presentation to control for possible rehearsal, it could happen even within a very short interval of time through an established habit. The reasoning that this type of rehearsal is through an established habit associated with RIGHT can be inferred from the results obtained in the AN-RW condition. In the AN-RW condition, even though S was told ARMY represented RIGHT, he had to make sure that ARMY represented RIGHT before safely rehearsing the SCHOOL-A-ARMY sequence. Otherwise, he might rehearse an incorrect sequence. This additional stage of seeking assurance or checking naturally decreased the time utilized for rehearsal. However, if an S in the AN-RW condition had a thorough practice, he might do as well as in the RW condition.

According to the foregoing analysis, Buchwald's (1969) model justly applied only to the AN condition in the present experiment. Under the AN condition, if correct responses are repeated more often than incorrect responses, it is because for some stimuli S recalls both his previous responses and their outcomes, and he repeats the responses when the outcome was a RIGHT indicator and does not when it was a WRONG indicator. There is no additional factor of rehearsal because S is not certain which sequence of events, SCHOOL-A-NAVY or SCHOOL-A-ARMY, to rehearse. Without rehearsal, it may take many trials before S can repeat correct responses more than expected by chance. This conjecture can be simply tested by increasing the number of information trials under the AN condition. Thus, Buchwald's model does not pay sufficient attention to the role of rehearsal. Depending only upon Buchwald's machinery, the usual fast rate of acquisition in a Thorndikian experiment cannot be accounted for.

The fact that no difference was obtained in WW repetitions among the various outcome conditions indicated that S rehearsed only when a correct sequence occurred. If he were to rehearse a correct sequence when he experienced an incorrect sequence, S would have to translate the incorrect sequence (SCHOOL-A-WRONG) into the correct sequence (SCHOOL-C-RIGHT). But it is unlikely that enough time was available to him because *E* was attempting to minimize the intervals between responses and successive stimulus presentations.

REFERENCES

BUCHWALD, A. M. Variations in the apparent effects of "right" and "wrong" on subsequent behavior: A new interpretation. *Journal of Verbal Learning and Verbal Behavior*, 1962, 1, 71–78.

BUCHWALD, A. M. Effects of "right" and "wrong" on subsequent behavior: A new interpretation. *Psychological Review*, 1969, 76, 132–143.

LIU, I., & CHUANG, C. Scaling 1,200 Chinese characters for meaningfulness. *Acta Psychologica Taiwanica*, 1970, 12, 33–52.

POSTMAN, L. Rewards and punishments in human learning. In L. Postman (Ed.), *Psychology in the making*. New York: Alfred A. Knopf, 1962.

(Received January 29, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 370–374

CRITERION EFFECTS IN SIMPLE REACTION TIME:
RESULTS WITH STIMULUS INTENSITY AND DURATION MANIPULATIONS[1]

A. J. SANFORD[2]

Medical Research Council, Applied Psychology Unit, Cambridge, England

It is shown that the proportion of loud to soft stimuli in a simple reaction time (RT) task affects both mean (overall) RT and the difference in RT to loud and soft stimuli. Such a finding can be readily interpreted in terms of a variable reaction criterion model. It is further demonstrated that analogous results can be obtained by varying loudness through stimulus duration instead of through amplitude. This both broadens the scope of possible accounts of the stimulus sampling mechanisms underlying the simple reaction process and provides a problem for models invoking sensory integration effects as the basis of stimulus evidence accumulation in simple RT.

Much recent research on simple reaction time (RT) to suprathreshold stimuli has been concerned with demonstrating that the effect of stimulus intensity on RT may be large or small depending upon the exact conditions of testing. There is, as a result, a considerable amount of evidence that the magnitude of the effect of loudness on simple RT depends upon overall adaptation conditions (Grice, 1968; Grice & Hunter, 1964), trial-to-trial sequential properties of the stimuli (John, 1967a) and temporal uncertainty (TU) (Kellas, Baumeister, & Wilcox, 1969). More recently Murray (1970) has demonstrated that incentive also determines the size of the effect of loudness on RT and that the frequency distribution of intensity levels in the stimulus series also has an effect. Thus, the loudness effect is greatest when there is a preponderance of loud stimuli, and it is smallest when there is a preponderance of quiet stimuli in the sequence.

The finding that the size of the loudness effect is so readily controllable implies that the locus of the mechanism giving rise to the loudness effect is central and not peripheral in origin, especially since preparatory state variables have an influence, as indicated above. Various 'counter models' have been put forward to account for the effect of loudness on RT (Grice, 1968; John, 1967b; LaBerge, 1962; McGill; 1963). There are three basic assumptions common to these models. First, it is assumed that evidence concerning the presence of a signal accumulates over time. Second, evidence accumulates at a faster rate for loud than for soft signals, and third, the amount of evidence required to elicit a response is not fixed but depends on experimental conditions. This leads to the prediction that where RT increases are due to changes in the amount of evidence required to elicit a response, RTs to both loud and soft stimuli will be increased, and there will be an increase in the difference between RTs to loud and soft stimuli.

This paper is concerned with two experimental attempts to further the basis of this type of model of RT. The first experiment involves a manipulation of the frequency distribution of intensity levels within a series, and is thus a replication, with slight differences, of Murray's (1970) study. The second experiment extends the argument and is an examination of the same paradigm applied to the manipulation of loudness through stimulus duration rather than amplitude. It is known (e.g., Port, 1963) that short bursts of noise sound quieter than long bursts of noise up to a limit of about 70 to 100 msec. even if they have the same physical amplitude. It is also known (Raab, 1962) that a 2-msec. pulse of white noise is responded to more slowly than a 40-msec. pulse, although the effect is quite small. Since the effect *is* so small there have been few, if any, attempts to replicate Raab's finding. In this study we report an attempt to replicate Raab's finding and examine the possibility of obtaining experimental variations of the stimulus duration-RT effect analogous to variations in the magnitude of the intensity-RT effect.

---

[1] This report is based on work carried out in partial fulfillment for the degree of PhD from the University of Cambridge. Thanks are due to D. E. Broadbent who supervised the work and to the Medical Research Council for support through a research studentship.

[2] Requests for reprints should be sent to A. J. Sanford, who is now at the Department of Psychology, University of Dundee, Dundee, DD1 4HN, Scotland.

EXPERIMENT I

In this study amplitude changes were used as stimuli.

Method.—The 24 *S*s (volunteer housewives) were randomly assigned to the two conditions of the experiment. The two conditions used were "loud probable" in which 75% of the stimuli were loud and "soft probable" in which 75% of the stimuli were soft. During the test session, 120 RTs were taken, comprising 30 of the rare stimulus and 90 of the probable stimulus. Stimuli were presented binaurally through headphones and consisted of increments in an ongoing background of random noise, the overall levels being 67 db. for the masking noise, rising to 72 db. for the quiet signal and 86 db. for the loud signal. Levels were measured by coupling the headphone to a sound-level meter. Rise times were effectively instantaneous, but electronic switching and the use of noise as the stimulus precluded any audible switching transients. Stimulus duration was 50 msec.

Foreperiod values of 1, 1.5, and 2 sec. were used, the imminence of a stimulus being indicated by a warning light of fixed intensity and duration .5 sec. Foreperiods were presented in random order, but each of the three values occurred on one-third of all occasions in each condition and for each intensity. Trials were presented once every 15 sec. and *S*s were given a brief rest after 60 trials. No knowledge of results was given until the end of the entire session. The *S*s were simply told to respond as quickly as possible, and that they would be told how they had performed at the end of the test. RTs were measured to the nearest csec. by a system of decatron timers from the stimulus onset to the closure of a normally open microswitch. The *S*s always used the forefinger of their preferred hand. Prior to the test session a brief practice session was given to acquaint *S*s with the stimuli which were used.

Results.—Fig. 1 and Table 1 show mean RTs averaged over all *S*s.

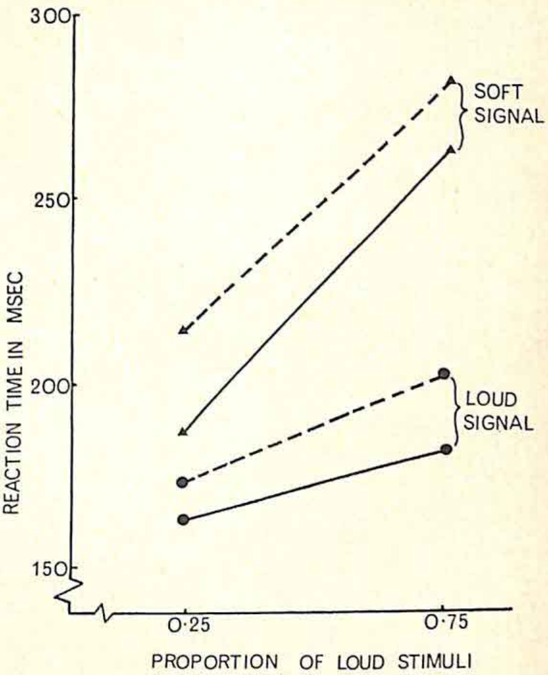


FIG. 1. The effect on RT to loud and soft signals as the proportion of stimuli is changed. (The groups are independent and the points are joined for clarity. Dotted lines, first half, full lines, second half.)

An analysis of variance was carried out using individual mean RTs. There were significant main effects of condition, $F\ (1, 22) = 7.85, p < .05$; intensity, $F\ (1, 22) = 54.38$, $p < .01$; and halves of test, $F\ (1, 22) = 6.33$, $p < .05$; on RT, and a significant Conditions × Intensity interaction, $F\ (1, 22) = 10.44$, $p < .01$. The Conditions × Halves of Test interaction was not significant, and no higher order interactions were significant. This is interpreted as follows: In the soft probable condition the size of the loudness effect is smaller than it is in the loud probable con-

TABLE 1

REACTION TIMES TO NEAREST MILLISECOND FOR EXPERIMENT I

Cond. (Group)	RT to loud		RT to soft		Difference	
	1st ½	2nd ½	1st ½	2nd ½	1st ½	2nd ½
Loud Probable	202	185	285	269	83	84
Soft Probable	174	163	216	187	42	24

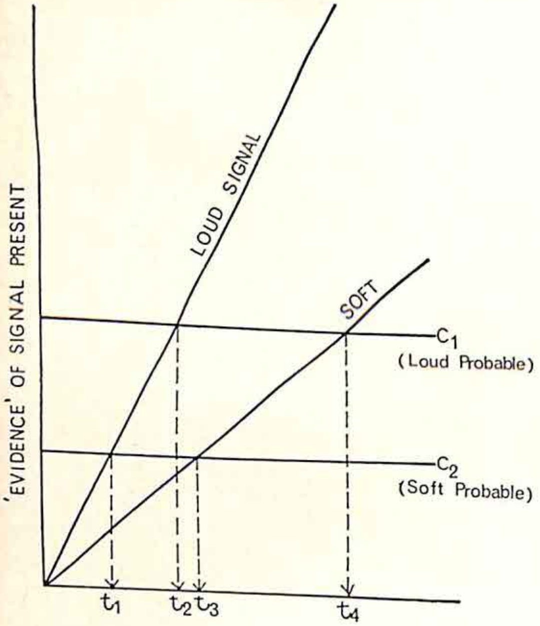


FIG. 2. The interpretation of the influence of loud or soft signal probability on RT and the size of the loudness effect. (The vertical axis is the amount of evidence after time t that a signal has been presented, where a signal is presented when $t = 0$. In accordance with the theory, evidence is accumulated faster for loud than for soft signals. C_1 and C_2 represent the amount of evidence required to elicit a reaction when the stimulus sequence is predominantly loud or soft, respectively. These criteria give rise to reaction initiations at times t_1, t_2, t_3, and t_4. Under predominantly loud sequences, the intensity effect on RT $(t_4 - t_2)$ is greater than it is under predominantly quiet sequences $(t_3 - t_1)$.

dition, and, correspondingly, mean RT is smaller in the soft probable condition.

Discussion.—This particular set of results is in complete agreement with the general findings of Murray (1970) cited earlier. As argued most fully by Murray, the most convenient descriptive interpretation is in terms of the criterion placement on the evidence accumulation functions assumed to underly simple RT as shown in Fig. 2. A preponderance of quiet signals appears to reduce the amount of evidence required to elicit a reaction in this situation. Why a preponderance of quiet signals should influence the criterion in this way is difficult to say, but one particularly appealing possibility is that it may be contingent on *S*s knowledge of his own RT. Certainly *S*s *judge* the slower RTs to soft signals as *being* slower (Sanford, 1970). In the present situation, it is therefore possible that in the soft probable case the criterion is lowered, because to keep it at a high level would involve a large number of unacceptably long RTs to the soft signal—i.e., a large number of unacceptably long RTs to the soft signal—i.e., a large number of unacceptably long RTs to the most dominant signal. In the case of the loud probable distribution, no such adjustment would be necessary since RTs to loud signals will often be of acceptable duration. The only real way to test this view would be to see whether *S*s maintain a fixed value of what is acceptably slow for RTs. This would require running the *same* *S*s in the different conditions of this experiment, and we are merely suggesting this as a possibility. The main aim of this paper is to show that effects obtainable through intensity manipulation can also be found through duration manipulation.

EXPERIMENT II

Method.—This is essentially a repeat of the Exp. I paradigm but with two differences in detail. Instead of using two intensities of fixed duration, intensity is fixed at 86 db. (background 30 db.), and two durations are used: 2 msec. (short) and 50 msec. (long). These combinations of background, intensity, and duration gave a reasonably good difference between RTs to long and short stimuli in a pilot investigation. Thirty-six *S*s were used (housewives from the Medical Research Council Applied Psychology Unit panel), being randomly allocated to three groups of 12. As in Exp. I, there were two groups, short probable (short 75%, of the time), long probable (long 75%), but an "equiprobable" group was also included. Prior to the test a brief practice session was given to acquaint *S*s with the stimuli which would be used.

In all respects of procedure, the method followed that used in Exp. I. RTs were measured from the stimulus onset to the closure of a normally open microswitch to the nearest csec.

Results.—Table 2 and Fig. 3 show a pattern of results similar to the previous experiment. A sequence containing pre-

TABLE 2

REACTION TIMES TO NEAREST MILLISECOND FOR EXPERIMENT II

Cond. (Group)	RT to long		RT to short		Difference	
	1st ½	2nd ½	1st ½	2nd ½	1st ½	2nd ½
75% long	213	192	301	258	88	66
50% long	187	180	230	216	43	36
25% long	176	171	220	197	44	26

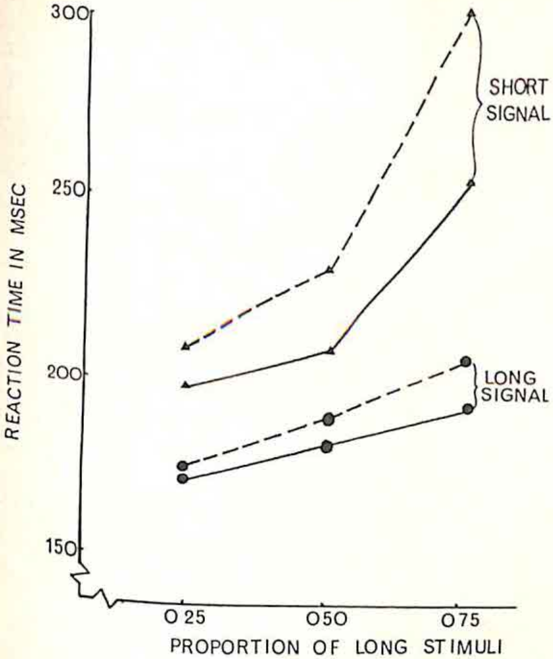


FIG. 3. The effect on RT to long and short signals as the proportion of stimuli is changed. (Groups independent, lines joined for clarity. Dotted lines, first half; full lines, second half.)

dominantly short (2 msec.) stimuli gives lower overall RTs than that containing predominantly long (50 msec.) stimuli. The sequence containing equal proportions is halfway between the two. An analysis of variance carried out on mean RTs showed that the effect of conditions was significant, $F\ (2,\ 33) = 4.03,\ p < .05$; duration had a significant effect, $F\ (1,\ 33) = 114.05,\ p < .01$; and practice (test half) was significant, $F\ (1, 33) = 18.21,\ p < .01$. The Conditions × Duration interaction was significant, $F\ (2,\ 33) = 8.15;\ p < .01$, which is again analogous to the result obtained in the intensity manipulation experiment. One further effect was significant, namely, the halves of the Test × Duration interaction, indicating that the size of the stimulus duration—RT effect depends on practice, $F\ (2,\ 33) = 11.92$; $p < .01$.

GENERAL DISCUSSION

Raab's (1962) finding that duration of a burst of noise influences RT has been replicated using only two durations although the effect here is much larger than the effect Raab obtained, especially in the first part of the experiment. A possible reason for this is that Raab's *S*s were very highly practiced (D. H. Raab, personal communication, 1971) while ours were unpracticed. More significantly, the size of the effect of duration on RT depends upon the make-up of the stimulus sequence in a manner analogous to the way the intensity-RT effect depends upon stimulus sequence, as we have demonstrated. In addition, since RT differences between loud and soft signals and RT differences between long and short signals is about the same in the two experiments, it would appear as though the effect of proportion of loud to soft stimuli and the effect of proportion of long to short stimuli is of the same order of magnitude, at least within the confines of the present stimulus values.

Although a wider range of manipulations of the RT-duration effect is desirable, it seems reasonable to suppose that there may be a common mechanism underlying the results of the two experiments performed above. If this is the case, and it appears to be, then the result carries implications for the nature of the stimulus evidence accumulation process. It is necessary to postulate that the rate of evidence accumulation of the models discussed in the introduction is a function of stimulus *duration* as well as stimulus amplitude. This is made necessary even by finding an effect of stimulus duration *at all* on RT, and is made even more necessary by finding an analogue of a criterion effect in the present experiment. Such considerations rule out one possible specific model of the evidence growth process. It would be parsimonious to suppose that the integration of energy within the critical duration of about 100 msec. represents the period during which evidence concerning a signal's presence is accumulated (the growth of evidence)—different intensity signals reaching different asymptotic levels after the critical duration. It is convenient then to suppose that it is during this period that the evidence accumulated passes the criterion. (Such a model has been hinted at by Teichner, 1954, and John, 1967b, both of whom cite energy integration effects as evidence of the "growth of sensation.") The present Exp. II indicates against this attractive theory, since the theory would predict no difference in the rate of evidence accumulation for stimuli of different durations provided they were of the same amplitude. Thus, if the evidence accumulation is part of the RT process, evidence accumulation cannot be determined by the energy integration process determining the "growth of loudness." We suggest that detailed models of the evidence accumulation process which is at the heart of current simple RT models must be able to account for stimulus duration effects as well as stimulus intensity effects. The presence of criterion-like effects in the size of the stimulus duration effect on RT should also be explored in more detail.

REFERENCES

GRICE, G. R. Stimulus intensity and response evocation. *Psychological Review*, 1968, **75**, 359–373.

GRICE, G. R., & HUNTER, J. J. Stimulus intensity effects depend upon the type of experimental design. *Psychological Review*, 1964, **71**, 247–256.

JOHN, I. D. Sequential effects of stimulus characteristics in a serial reaction-time task. *Australian Journal of Psychology*, 1967, **19**, 35–40. (a)

JOHN, I. D. A statistical decision theory of simple reaction-time. *Australian Journal of Psychology*, 1967, **19**, 27–34. (b)

KELLAS, G., BAUMEISTER, A. A., & WILCOX, S. J. Interactive effects of preparatory intervals, stimulus intensity, and experimental design on reaction-time. *Journal of Experimental Psychology*, 1969, **80**, 311–316.

LABERGE, D. A recruitment theory of simple behaviour. *Psychometrika*, 1962, **27**, 375–396.

MCGILL, W. J. Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, & E. Gallanter (Eds.), *Handbook of mathematical psychology*. Vol. 1. New York: Wiley, 1963.

MURRAY, H. G. Stimulus intensity and reaction time: Evaluation of a decision-theory model. *Journal of Experimental Psychology*, 1970, **84**, 383–391.

PORT, E. Über die lautstarke einzelner kurzer Schallimpulse. *Acustica*, 1963, **13**, 212–223.

RAAB, D. H. Effect of stimulus duration on auditory reaction time. *American Journal of Psychology*, 1962, **75**, 298–301.

SANFORD, A. J. Rating the speed of a simple reaction. *Psychonomic Science*, 1970, **21**, 333–334.

TEICHNER, W. H. Recent studies of simple reaction time. *Psychological Bulletin*, 1954, **51**, 128–149.

(Received February 1, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 375–381

TRANSFER AS A FUNCTION OF STIMULUS, RESPONSE, AND SIMULTANEOUS STIMULUS AND RESPONSE SIMILARITY

BARBARA S. UEHLING[1]
Roger Williams College

BENTON J. UNDERWOOD
Northwestern University

Three experiments were conducted in which both early and late transfer measures were compared as a function of interlist similarity of stimuli, of responses, or of both stimuli and responses in two successive paired-associate lists. Three degrees of similarity were employed within each experiment. Trial 1 transfer was found to increase from negative to positive when stimuli varied in similarity and responses were identical, to increase from negative to zero when responses varied in similarity and stimuli were identical, and to remain about zero as similarity of stimuli and responses varied simultaneously. Trials to criterion transfer resulted in no significant effect of similarity. However, interlist response variation produced inferior performance to that obtained with stimulus variation or both stimulus and response variation. Results were interpreted within a mediation framework.

The purpose of this study was to examine the influence of simultaneous variation of interlist semantic stimulus similarity and semantic response similarity on paired-associate transfer. For comparative purposes, stimulus similarity and response similarity were also manipulated independently. These latter manipulations have occurred in a number of experiments and their effects are reasonably consistent for the initial transfer trials. With response words in both lists identical, performance on the transfer list increases as stimulus similarity increases from zero to a high level (e.g., Kasschau & Pollio, 1967). With stimulus words identical, transfer performance also increases as the response terms increase in similarity (e.g., Postman, 1966).

Evidence on the simultaneous variation of stimulus and response similarity for verbal units is scanty. Wimer's data (1964) indicated that positive transfer occurs when both stimulus and response terms are similar, but the magnitude of the effect was small and the nature of the similarity involved is difficult to specify. Shea (1969) has shown some increase in transfer performance as stimulus similarity and response similarity increased. However, since Shea used a single multiple-choice transfer test there was no evidence on the transfer effects over trials on the second list. Transfer effects as a function of similarity may differ appreciably depending upon the stage of learning of the transfer task at which they are observed, and they depend also upon the amount of experience *S* has had with a given paradigm (Postman, 1964). Indeed, such evidence makes it difficult to specify a set of laws relating similarity and transfer (as in a transfer surface) which has any degree of generality. Theoretical interpretations (see Discussion section) obviously must account for these shifting transfer effects.

METHOD

General procedure.—Three parallel transfer experiments were conducted. In all three experiments, similarity in terms of degree of synonymity of adjectives was manipulated, but the experiments differed in which member of the pairs varied in similarity. In Exp. SV, stimulus similarity was varied while the response terms remained identical; in Exp. RV, response similarity was varied with the stimulus terms identical; and, in Exp. SV-RV, similarity of both stimuli and responses was varied simultaneously.

Three conditions of similarity were investigated within each experiment; low (L), medium (M), and high (H). Each *S* served in all similarity conditions within an experiment, but independent groups of *S*s served for each of the experiments. The experiments and similarity conditions within each are summarized in Table 1.

Subjects.—A total of 72 Northwestern University undergraduate students served as *S*s, 24 in each experiment. All were volunteers from introductory

---

[1] Requests for reprints should be sent to Barbara Uehling, Roger Williams College, Bristol, Rhode Island 02809.

TABLE 1

Experimental Conditions

Exp.	Similarity		
	High	Medium	Low
SV	A–B, A′–B	A–B, A″–B	A–B, C–B
RV	A–B, A –B′	A–B, A –B″	A–B, A–C
SV–RV	A–B, A′–B′	A–B, A″–B″	A–B, C–D

psychology classes and none had previously served in a study requiring paired-associate learning.

Materials.—Each list consisted of eight pairs of adjectives taken from Haagen (1949). All items used were roughly equivalent in familiarity (.5–.8, according to Haagen). Items chosen for the H conditions were rated between .8 and 1.2 on Haagen's similarity of meaningfulness scale; those chosen for the M conditions had ratings of 2.3 to 2.7, while the L condition was made up of items appearing to be unrelated.

Two sets of transfer lists were prepared for each degree of similarity used in Exp. SV and RV. These sets were then combined into one set for Exp. SV-RV. Two sets were needed because the simultaneous variation of similarity in Exp. SV-RV required one set of similar items for stimuli and another for responses. Half of the *S*s in Exp. SV and RV received the Set 1 lists, and half, the Set 2 lists. An illustration of the alterations in pair arrangements in the various experimental conditions appears useful. Two pairs, one from each set, from the H condition of Exp. SV were WICKED-UNTAUGHT and TRANQUIL-IMMENSE. The corresponding transfer pairs were EVIL-UNTAUGHT and PEACEFUL-IMMENSE. Members within the pairs were simply reversed to form the analogous pairs for the RV experiment. Hence, the first pairs became UNTAUGHT-WICKED and IMMENSE-TRANQUIL, while their corresponding transfer pairs were UNTAUGHT-EVIL and IMMENSE-PEACEFUL. In the SV-RV experiment, the pairs from both sets were combined to provide the simultaneous stimulus and response variation; thus one transfer list contained a pair combining the just-cited pairs to form TRANQUIL-WICKED, while the corresponding transfer pair was PEACEFUL-EVIL.

Procedure.—All *S*s in all experiments served for 4 successive days. The first day was a practice day in which two lists of paired associates were learned. The same two practice lists were given to all *S*s. Moderate similarity of both stimuli and responses was used, thus corresponding to the M condition of Exp. SV-RV. The *S*s were assigned to one of the three experiments on the basis of number of trials required to learn the practice list. On each of the 3 remaining experimental days, one of the three similarity conditions within the assigned experiment was presented. Order of similarity presentation was counterbalanced across *S*s, as was the order of two lists within a transfer condition.

All lists were presented at a 2:2-sec rate on a Hull-type memory drum with an intertrial interval of 8 sec. All lists were learned to a criterion of 8/8 or a minimum of 5 trials for the second list. Three orders of the pairs were used in all conditions. Learning of the second list within each session always followed immediately after learning of the first list.

Results

Practice day.—Assignment of *S*s to the three experiments resulted in equivalent groups. The mean number of trials to attain the criterion of 8/8 on the first practice list was 16.58, $\sigma_M = 1.51$; 16.71, $\sigma_M = 1.38$; and 16.58, $\sigma_M = 1.37$ for Exp. SV, RV, and SV-RV, respectively. The largest difference between experiments resulted in a t of only .05. Variability of scores was also comparable according to a Hartley test for homogeneity of variance.

List 1 learning.—Differences in transfer among experiments should not be confused with differences in list difficulty. Equality of list difficulty can be evaluated by a comparison of trials to learn the first list in all conditions since every list was presented as a first list for half of the *S*s. The range of means among the nine conditions was from 12.79 to 15.13. An analysis of variance of these data resulted in an $F < 1.00$. List 2 results can therefore be directly compared.

Transfer results.—Analyses of variance were conducted for both (*a*) early transfer, i.e., number of correct responses on Trial 1 of List 2 learning and for (*b*) later transfer, i.e., number of trials to criterion for List 2 learning. These analyses were conducted for all three experiments simultaneously. However, before the analyses incorporating data from all experiments could be done, it was necessary to establish that neither the source of variance contributed by the two sets of words used in Exp. SV and RV nor their order of presentation was significant; neither one was found to be significant.

The mean number of correct responses given on Trial 1 of List 2 for all conditions is given in the upper panel of Fig. 1. The units on the abscissa are represented as equal, although this is obviously an assumption since there are no scale values for the materials used in the L conditions. The analysis of variance conducted on Trial 1 transfer resulted in two significant sources

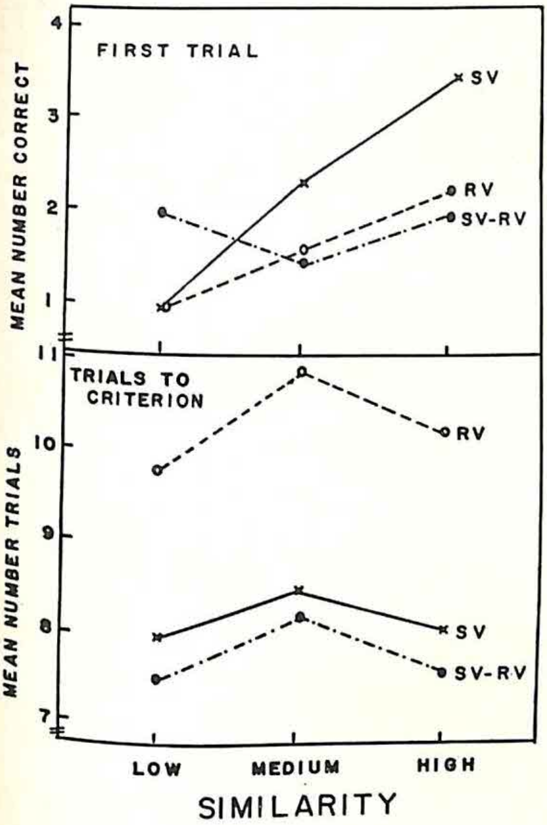


FIG. 1. First trial and trials-to-criterion transfer as a function of similarity in Exp. SV (stimulus variation), RV (response variation), and SV-RV (simultaneous stimulus and response variation).

of variance: similarity, F (2, 132) = 14.26, and the interaction of similarity with experiments, F (4, 132) = 4.84, both $p < .01$.

The most appropriate comparison group for evaluating similarity effects is the L condition of SV-RV (A-B, C-D) since learning-to-learn and warm-up effects for this condition should be the same as for the others, but the items are unrelated in similarity. A Duncan multiple-range test resulted in significant differences ($p < .05$) between the low-similarity conditions of both SV and RV (A-B, C-B; A-B, A-C) and the A-B, C-D condition, indicating negative transfer for both of these conditions. Performance in the H condition of the SV experiment (A-B, A′-B) was also significantly greater ($p < .05$) than for all other conditions. Transfer in the H condition of the RV experiment (A-B, A-B′) was essentially zero since it did not differ significantly from the A-B, C-D condition. However, the H condition of RV did differ significantly from the L condition of that experiment. Magnitude of similarity effects upon early transfer is greater with SV than RV, ranging from negative to positive for stimulus variation of similarity, while ranging from equivalently negative transfer to zero transfer for response variation.

None of the conditions of the SV-RV experiment differed from each other, indicating essentially zero transfer for all conditions.

The mean numbers of trials to attain a criterion of 8/8 are shown in Fig. 1. These data are quite different from those of Trial 1. An analysis of variance showed no significant effect of similarity, F (2, 132) = 1.08. However, the difference among experiments was significant, F (2, 54) = 5.64, $p < .01$. Orthogonal comparisons resulted in no significant difference between the SV and SV-RV experiments on the trials-to-criterion measure. However, a comparison between Exp. RV and the other two experiments approached conventional levels of significance, F (1, 69) = 3.71, $p < .07$.

In order to examine more closely where during List 2 learning, the diminution of similarity effects on transfer occurs, trials to successive criteria were plotted for all conditions. From these figures it was found that the relative similarity effects among conditions which were displayed on first-trial transfer remained until late in second-list learning, i.e., acquisition of the sixth and seventh items. The difference between transfer results for early transfer and trials-to-criterion transfer can be accounted for by changes occurring late in learning rather than during the middle portion.

Intrusions from List 1 which occurred in each third of learning List 2 are shown in Table 2. Trials to attain the criterion of List 2 learning for each S were divided into thirds. In the case of a number of trials not divisible by 3 without a remainder, division was made differently for the remainder 1 or 2. If 1, rounding to the fewer number of trials, was used as the basis for the first and final thirds of learning but

TABLE 2

Intrusions during List 2 Learning

Exp. and degree of similarity	Stage of List 2 learning		
	First third	Middle third	Final third
SV			
High	2	0	0
Medium	1	1	0
Low	0	0	0
RV			
High	18	10	4
Medium	5	14	8
Low	0	0	0
SV–RV			
High	15	3	2
Medium	1	1	0
Low	0	0	0

if the remainder was 2, rounding to the greater number of trials was used as the basis for the first and final thirds of learning. As would be expected, few intrusions occurred in Exp. SV since an intrusion in this paradigm consisted of giving a List 1 stimulus as a response in List 2.

The total number of intrusions is greatest in the RV experiment with a different pattern emerging for H and M conditions, however. Intrusion frequency decreases as learning progresses in the H condition, but first increases then decreases in the M condition. Intrusions in Exp. SV-RV are most comparable to those in Exp. RV, at least for the H condition.

It is possible that transfer results of the present experiment were altered by using a within-*S*s design for similarity within an experiment. Postman (1964) has found that early transfer results will change in the A′-B paradigm (SV-H of this experiment) as a function of successive practice with this paradigm, amount of positive transfer increasing as a function of practice. He also found that negative transfer will decrease in the A-C paradigm (RV-L of this experiment). The transfer gradient in the present SV experiment might then be expected to become greater over experimental days as *S* has more experience with paradigms in which responses are identical but stimuli can be described as having some degree of similarity. Likewise, the transfer gradient when responses were varied could be expected to become attenuated over days since negative transfer effects with response variation would be expected to decline.

The initial practice day in which *S*s learned lists which fitted the M condition of SV-RV makes the prediction somewhat less clear. However, transfer gradients both for early and late measures were examined as a function of the 3 experimental days. Thus, the data for the eight *S*s who were given the H condition of SV on the first experimental day were compared with data for those receiving the SV-M condition and the eight receiving the SV-L condition

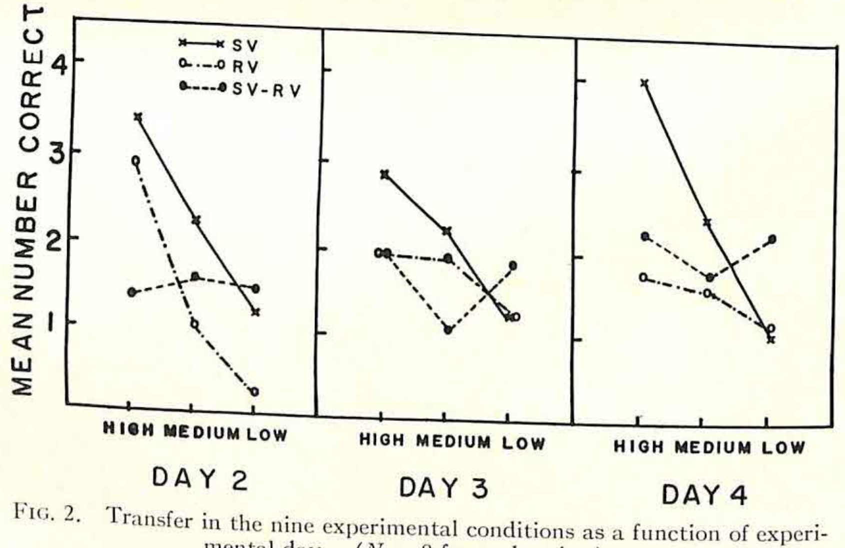


Fig. 2. Transfer in the nine experimental conditions as a function of experimental day. (N = 8 for each point.)

on that day. Results were likewise examined for the RV and SV-RV experiments. The remaining 2 experimental days were similarily examined for all three experiments. These results for first-trial transfer are shown in Fig. 2, in which it can be seen that the RV gradient does decline over days while the gradient for SV is greatest on the third day. The SV-RV data show no consistent changes from day to day and no clear gradients on any day. Results of the trials-to-criterion response measure when looked at in this way did not show any systematic trends and hence are not reported. These findings are consistent with Postman (1964) since he found that alterations which occur in transfer results due to the increasing experience with the same paradigm were restricted to early transfer measures.

DISCUSSION

The present study was concerned with early and late transfer effects of similarity manipulation of stimuli, of responses, and of simultaneous stimulus and response variation when compared with an A-B, C-D control. The results indicate that in early transfer, stimulus similarity variation has the greatest effect when transfer ranges from negative transfer when stimuli are dissimilar and responses identical (A-B, C-B) to high positive transfer with highly similar stimuli and identical responses (A-B, A′-B). Response variation produced a flatter gradient with about the same degree of negative transfer occurring in the L condition (A-B, A-C) as in the condition of stimulus variation. Negative transfer decreased with increasing response similarity until zero transfer was attained in the H (A-B, A-B′) condition. The relatively greater effect of SV than RV may be somewhat enhanced by the within-*S*s design since experience with stimulus variation has some tendency to increase the slope of the gradient, while experience with response variation seems to have the effect of decreasing the gradient slope.

It is an interesting question as to why these changes with experience occur. It would seem that *S*s with increasing experience alter the way in which they employ mediating associations since any parasitic effects of generalization based on similarity would be automatic and hence not readily modifiable by experience. Perhaps *S*s in the RV paradigm simply abandon the use of mediating similarity effects in learning List 2 as a function of increasing practice with the paradigm. To provide further information on this possibility, intrusions for the RV experiment were examined for each similarity condition as a function of successive experimental days. Although only eight *S*s were represented at each point, it is interesting to note that for the first experimental day, intrusions were 14, 6, and 0 for the H, M, and L conditions, respectively. For the second experimental day, the corresponding numbers were 13, 1 and 0, while for the third experimental day the H similarity condition resulted in 5 intrusions; the M condition, 20; and the L, 0. Rather than simply abandoning similarity as a mediating device, it would appear that *S*s are influenced in their recognition of similarity relationships by their prior experiences with the paradigm. If they have not had the H similarity condition until the last experimental day, they apparently use similarity less since only 5 intrusions occurred for those *S*s with this experience. However, if the H condition is given either on the first or second experimental day and the M condition on the final experimental day, *S*s appear to attempt to use a similarity relationship since 20 intrusions occur for *S*s with this experience.

Simultaneous stimulus and response variation transfer does not appear to be a joint function of independent stimulus and response manipulation, since nearly zero transfer results occur in all SV-RV conditions. These findings are not in accordance with Wimer (1964) or Shea (1969), who found some positive transfer with increasing similarity of both stimuli and responses. Early transfer in the SV-RV paradigm of the present experiment appears to more nearly resemble response variation of similarity than stimulus variation.

Late transfer results indicate no effect of similarity manipulations in any of the three experiments, but do indicate negative transfer for all conditions of Exp. RV. Thus, in late transfer measures, simultaneous stimulus and response variation more nearly resemble SV results with essentially zero transfer produced for all conditions.

These findings are most consistent with a mediation interpretation of transfer. In the SV experiment, early transfer should be positive in the H condition and to a lesser extent in the M condition, in which chaining from A′ to A to B or A″ to A to B would occur. These were the results and, as further support, two intrusions did occur in which A′ elicited A early in List 2 learning.

However, according to the mediation interpretation, as List 2 learning progresses and A and A′ become about equal in strength transfer should decrease. Since the mediated chain always results in B as a response to either A′ or A or A″ or A, interference might not be detected if enough time is permitted *S* to respond. However, latency measures might reflect a competition between A and A′ (A and A″) or the failure to discriminate between mediator and mediated, as Postman (1964) has described this competition. Relatively less assistance from an A′-A chain should occur as List 2 learning progresses for those associations which are weak at the beginning of List 2. The present data indicate a decline in positive effects during List 2 learning so that no significant difference is detectable in the total transfer measure.

In the A-B, C-B condition, mediated positive transfer effects would not be expected, nor are they present. Instead, negative transfer would be expected to result, since after learning the association A-B the chain C-A-B might be needed in order to meet the requirement of learning C-B. Interference would also be present to the extent that backward learning of A-B has occurred and is no longer appropriate. Whether *S* is learning C-A-B or extinguishing B-A or both, as List 2 learning progresses and C becomes relatively stronger than A, negative transfer should decrease. Results of the present study are consistent with this expectation.

A mediation interpretation of response similarity variation in transfer would imply that A-B′ or A-B″ following learning of A-B will be via the chain A-B-B′ or A-B-B″. This chaining will be facilitated initially to the extent that B is associated with B′ and B″ by preexperimental associations and to the extent that strong A-B learning has been built in. The present data are consistent with this interpretation. During subsequent List 2 learning, however, as B and B′ become equal in strength and hence competitive, less positive transfer and perhaps negative transfer would be expected. This effect is reflected in the total transfer measure. Intrusion data also support this interpretation since the greatest number of intrusions occurred in the A-B′ condition and most of these were early in learning. That the number of intrusions was greater for the middle third of learning than the initial third for the A-B″ condition suggests that B and B″ become equal in strength later in learning than B and B′, in which the learning process is accelerated by greater preexperimental similarity effects.

The A-C condition should result in negative transfer throughout learning since there are no initial positive effects of mediated similarity and the chaining A-B-C would produce interference. Again results are in line with this interpretation.

Extending the mediation hypothesis to the simultaneous variation of stimulus and response would suggest that for the H condition, the chaining mechanism would be A′-A-B-B′. Although the chain is somewhat longer and hence latencies might be longer early in List 2 learning, the effect should be to produce a weak positive transfer early in transfer, at least for strong A-B associations. This would be true because of the combination of positive strength from A′-A-B, which would occur with stimulus variation, and A-B-B′, which would occur with response variation. It is possible that weak A-B associations would lead to interference for an *S* who is forced to respond in a short interval of time. As learning progresses, failure of the distinction between A and A′ as well as the increase in strength of B′, and hence competitive strength with B, should lead to a reduction in transfer. Moderate similarity of both stimuli and responses would be expected to produce less positive transfer because mediated effects would be less than for the high-similarity condition. Although the present data are not inconsistent with this interpretation, the relative effects appear to be very weak. Latency measures for responding might be more sensitive to such facilitation and competition effects.

Since both stimuli and responses are different in the SV-RV H and M conditions, it might be thought that *S* simply ignores the similarity relationship and proceeds to learn the list as though it is an A-B, C-D list. However, the intrusion data would argue against this interpretation since there are nearly as many intrusions in the SV-RV H condition as the RV-H condition, suggesting that the chaining mechanism is operating but that the negative effect of learning a new response B′ is interfering, and is not offset by the A′-A-B facilitation. This is consistent with the greater apparent similarity of the SV-RV results to RV in early transfer, and to an interpretation that response variation is more critical in determining early transfer when both stimuli and responses are manipulated in similarity than is stimulus variation.

Using mediation, the gradient of transfer which should be obtained for SV should be steeper than for RV in early transfer, at least with meaningful materials where response learning is minimized (Jung, 1963). In the A-B, A′-B paradigm positive transfer should occur with the chain A′-A-B, while in the A-B, C-B condition negative transfer should occur with the chain C-A-B required by the learning task. Thus, conditions range from high positive transfer to negative.

With response variation of similarity, however, maximum similarity should yield only slight positive transfer with the chain A-B-B′, but the dissimilar condition A-C will require learning the chain A-B-C and hence will produce negative transfer.

For later transfer, positive effects of mediated association for both A′-A-B and A-B-B′ should decrease when the similar items become about equal in strength. However, since either A or A′ will produce the coerrect reponse, interference should not result. With B and B′ in competition, however, where only one is correct, interference should result. Thus as occurred in this experiment, early positive transfer effects due to mediated similarity should disappear and only interfering effects of competing responses should remain, until List 2 items become relatively stronger. That *S*s can adopt a strategy which utilizes these positive effects but decreases negative effects with continued exposure to the paradigm was found by Postman (1964) and is suggested to some extent by the present data. Practice with a paradigm manipulating both stimulus and response similarity might, according to the mediation interpretation, result in increasing positive transfer since *S* might be able to use stimulus similarity relationships which should increase transfer with practice and lessen the negative transfer effects of response dissimilarity.

REFERENCES

HAAGEN, C. H. Synonymity, vividness, familiarity, and association-value ratings of 400 pairs of common adjectives. *Journal of Psychology*, 1949, **27**, 453–463.

JUNG, J. Effects of response meaningfulness (m) on transfer of training under two different paradigms. *Journal of Experimental Psychology*, 1963, **65**, 377–384.

KASSACHAU, R., & POLLIO, H. Response transfer mediated by meaningfully similar and associated stimuli using a separate-lists design. *Journal of Experimental Psychology*, 1967, **74**, 146–148.

POSTMAN, L. Studies of learning-to-learn. II. Changes in transfer as a function of practice. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 437–447.

POSTMAN, L. Differences between unmixed and mixed transfer designs as a function of paradigm. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 240–248.

SHEA, M. Formulation of a generalization surface for the simultaneous variation of stimulus and response similarity. *Journal of Experimental Psychology*, 1969, **80**, 353–358.

WIMER, R. Osgood's transfer surface; extension and test. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 274–279.

(Received February 2, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 382–387

FACILITATION OF HUMAN OPERANT RESPONDING BY STIMULI WHICH PRECEDE AVERSIVE EVENTS[1]

RODNEY A. POETTER AND PAUL LEWIS[2]

Ohio University

The effects of stimuli preceding timeout, response cost, and point-loss periods on the operant response rate of human *S*s were investigated. The *S*s' response rates during a stimulus light preceding the onset of the aversive event, during the aversive event, and immediately following termination of the aversive event were compared with *S*s' base-line response rates. The results indicate that (*a*) response-rate acceleration is characteristic of *S*s' response patterns during the preaversive stimulus for all three experimental conditions, (*b*) the response-acceleration effect can be demonstrated to schedules of reinforcement with very short intervals between reinforcement delivery—hence, rapid rates of responding, (*c*) increased reinforcement density is not a necessary condition for the occurrence of the facilitation effect, (*d*) noncontingent punishment is ineffective in suppressing response rate, and (*e*) there is no evidence of post-aversive stimulus acceleration.

The term "response cost" refers to the cost to an organism of making a response. More specifically, response cost is the removal, contingent upon a response, of a positive reinforcer that is concurrently, or previously was, being employed to maintain that behavior. Thus when response cost is applied to a response, that response is maintained and reinforced by the presentation of a positive reinforcer and *concurrently* "punished" or followed by response-contingent withdrawal of that same reinforcer.

Timeout, as well as response cost, involves a loss to the organism; however, instead of the removal of positive reinforcement already received by *S*, timeout represents the loss of the opportunity to earn further reinforcement. More specifically, timeout is defined as "a period of nonreinforcement arranged by extinction in the presence of a particular stimulus or by removal of an opportunity to respond . . . [Catania, 1968, p. 348]."

An aversive stimulus may be defined as having two general characteristics: (*a*) when made contingent upon a response, an aversive stimulus reduces the frequency or probability of that response, and (*b*) the termination or avoidance of an aversive stimulus results in an increase in the frequency or probability of the escape or avoidance response in a similar stimulus situation. Given the above criteria, both response cost and timeout meet the definition of an aversive stimulus.

Several studies (Baron, Kaufman, & Stauber, 1969; Weiner, 1962, 1964a, 1964b, 1967, 1969a) have investigated the effects of response cost on performances maintained by various schedules of reinforcement and extinction, and the results consistently have shown suppression of responding. For example, Weiner (1962) employed a variable-interval (VI) schedule of reinforcement, which delivers reinforcement contingent on the first response after varying periods of elapsed time, to obtain a stable response rate; then, he imposed a cost condition, whereby 1 point was subtracted for each response, although points as reinforcers were delivered concurrently (100 points per reinforcement) to maintain responding. Response rate was suppressed markedly. Weiner (1963, 1969b) demonstrated both escape and avoidance responding by employing a point-loss period as the aversive event, i.e., a period of time during which points were subtracted unless the appropriate response was emitted.

---

[1] This research represents partial fulfillment of requirements for the MS degree, and was conducted by the first author under the direction of the second author. Thanks are due to Leonard W. Ferguson for his assistance with the manuscript preparation.

[2] Requests for reprints should be sent to Paul Lewis, Department of Psychology, Porter Hall, Ohio University, Athens, Ohio 45701.

Similarly, timeout has been shown to effectively punish responding and to reinforce escape and avoidance responding (Adelman & Maatsch, 1956; Baer, 1962; Ferster, 1957; Holz, Azrin, & Ayllon, 1963; Kaufman & Baron, 1968; McMillan, 1957; Morse & Herrnstein, 1956; Nigro, 1966; Willoughby, 1969); thus, when Baer applied response-contingent timeout to nursery school age children by interrupting film cartoons whenever undesirable thumbsucking occurred, a significant suppression of thumbsucking behavior resulted. And Morse and Herrnstein have demonstrated that pigeons learn to peck one key to avoid removal of a stimulus correlated with a schedule of positive reinforcement (food) on a second key.

Estes and Skinner (1941) originally advanced the procedure now commonly called the conditioned emotional-response (CER) paradigm. In the procedure, a stable base-line response rate is maintained by some schedule of positive reinforcement; then, a signal, or conditioned stimulus (CS), is presented for a predetermined time interval, followed by unavoidable shock. After several pairings of CS and shock, the result is a decrease in response rate during the CS.

Can one generalize from this finding with electric shock and state that response suppression occurs during a stimulus preceding other aversive events? The answer is no. When timeout is employed, response *acceleration* is generally observed during the pretimeout stimulus (Ferster, 1958; Leitenberg, 1966; Trenholme, Baron, & Kaufman, 1969), which is directly opposite the results obtained with shock. Trenholme et al. (1969) employed college students, with money serving as the positive reinforcer. After several 50-min. base-line sessions with a VI 1.33-min. schedule of reinforcement, a blue light was presented for 30 sec. followed by a 5-min. timeout period. Data were presented in the form of a ratio, i.e., number of responses during the CS/number of responses during an equal time period immediately preceding onset of the CS. The investigators reported a significant response acceleration, i.e., ratio greater than one, during the blue light.

The present experiment compares the rate of responding during a CS with the base-line rate when the CS is followed by timeout, response cost, or a point-loss period.

METHOD

Subjects

Fifteen female college students, enrolled in an introductory psychology course, served as *S*s. The *S*s were randomly assigned to one of the three experimental conditions: (*a*) timeout (TO), (*b*) response cost (RC), and (*c*) unavoidable point loss (UPL).

Apparatus

A 15 × 20 in. vertical panel was mounted on a table in *S*'s experimental room. Mounted on this operant panel, which was constructed from varnished plywood covered and reinforced with clear Plexiglas, were the following: (*a*) one response manipulandum (Gerbrands rat lever), (*b*) two evenly spaced 4-digit Sodeco counters located 9 in. from the bottom of the panel, one labeled "points gained" and the other "points lost," and (*c*) three evenly spaced lights located 2 in. from the top of the display, the left blue, the middle white, and the right yellow. A fan was employed to mask extraneous noise.

Standard electromechanical programming equipment located in an adjacent room delivered reinforcement (points) according to a VI schedule of reinforcement. Light onset and other stimulus conditions were controlled manually by *E*. Digital counters recorded stimulus and response events, and a Gerbrands cumulative recorder displayed *S*'s rates and patterns of responding.

Procedure

Instructions.—For each of the 15 *S*s the instructions contained the following information: (*a*) the session would begin when the white light came on and would end 50 min. later, (*b*) the task was to earn as many points as *S* could by pressing the bar, and (*c*) for every 50 points earned, *S* would receive a new "Snoopy cartoon," i.e., a recent *Peanuts* comic strip which *S*s received upon completion of each experimental session.

Prior to Sessions 5–10 for each *S* in the TO group, the sentence "when the yellow light is on you will not gain points" was included to facilitate discrimination between base-line and timeout periods (although *S*s were not instructed to *not* respond during timeout). Prior to all sessions for each *S* in the RC and UPL groups, *E* instructed *S* that "the number of points you earn will be determined by the number of points on the "points-gained" counter minus the number of points on the "points-lost" counter."

Sessions 1 and 2.—Pilot work demonstrated that human *S*s emit a stable response rate on a VI schedule of positive reinforcement in very few sessions, and they maintain responding for entire 50-min. sessions with points as reinforcers; hence, two sessions were sufficient to allow for the establishment of a fairly stable base line and adaptation to the novel stimulus situation.

During all sessions, a VI 10-sec. schedule of reinforcement delivered 1 point on the "points-gained" counter for each reinforcement; i.e., points were delivered contingent upon the first response following varying time intervals, the average of which was 10 sec. A fifty-interval VI tape was generated, with the duration of intervals employed ranging randomly from 1 sec. to 49 sec. (Fleschler & Hoffman, 1962).

Sessions 3 and 4.—The blue light (which served as the preaversive stimulus for all groups, was presented five times per session at random intervals, with the stipulation that at least 4 min. elapse between consecutive blue-light presentations to allow for stabilization of response rate. Immediately after termination of the blue light, the yellow light was presented. Duration of both lights was 30 sec. Timeout, response cost, and unavoidable point-loss periods were not presented during Sessions 3 and 4, for these sessions were necessary to eliminate any possible explanation of the mere presentation of the stimulus lights alone as the determining factor in a change of response rate.

Sessions 5–10.—The general procedure employed during Sessions 5–10 was the same as that during Sessions 3 and 4; however, during the yellow light the aversive event was presented. For the TO group, this consisted of a 30-sec. period of time during which responses were not followed by reinforcement. The responses of *S*s in the RC group were reinforced concurrently according to the VI 10-sec. schedule of reinforcement and punished by .1-point cost, or loss, registered on the "points-lost" counter in tenths of a point following each response during the yellow light. During the yellow light for *S*s in the UPL group, responses continued to be reinforced according to the VI schedule of reinforcement. Concurrently, points accumulated on the "points-lost" counter at the rate of .1 point/sec, i.e., a total loss of 3 points during each presentation of the yellow light.

Response Measures

Four response measures were obtained: (*a*) A—the number of responses during the 30-sec. period immediately preceding onset of the preaversive stimulus (blue light), i.e., base-line response rate, (*b*) B—the number of responses during the 30-sec. presentation of the preaversive stimulus, (*c*) C—the number of responses during the 30-sec. presentation of the aversive event, and (*d*) D—the number of responses during the 30-sec. period immediately following termination of the aversive event.

Three ratios were formed from the raw data: (*a*) B/A, (*b*) C/A, and (*c*) D/A. The B/A ratio, which is of primary interest, indicates the effect of the preaversive stimulus on the ongoing operant base-line response rate. A ratio greater than 1.0 indicates response acceleration; equal to 1.0, no effect; and less than 1.0, conditioned suppression. The relative rate of responding during the aversive condition is reflected in the C/A ratio. The D/A ratio indicates how the delivery of the aversive stimulus affected the base-line response rate immediately following termination of that aversive event.

RESULTS

Statistical analysis consisted of a three-factor mixed design, with the number of responses serving as the dependent variable, and a two-factor mixed design, with the particular ratio being investigated serving as the dependent variable (Bruning & Kintz, 1968). The former analysis controlled for the possibility that response acceleration could be more easily demonstrated with *S*s who emit a very low response rate; the latter controlled for wide discrepancies in base-line response rate between *S*s.

Because every experimental effect that was significant in one analysis was significant in the other, only the data concerning the ratios are presented. Post hoc analyses of significant main effects (experimental condition or session) and interaction effects (Group × Session) were conducted by employing the Newman-Keuls multiple-comparison statistic (Winer, 1962).

Base-Line Training: Session 1 and 2

A reasonably stable response rate was characteristic of all *S*s' response patterns by the end of Session 2. Number of responses emitted per 50-min. session was fairly consistent within *S*s, but varied greatly between *S*s—ranging from a low of 369 to a high of 14,990 responses.

Light-Alone Condition: Session 3 and 4

Appropriate tests (t) for related measures (Bruning & Kintz, 1968) revealed no significant difference between the number of responses during the blue-light, yellow-light, or post-yellow-light conditions, and the number of responses during an equal time period preceding onset of the blue light. This analysis indicates that the presentation of the stimulus lights alone does not affect significantly the base-line response rate.

Effects of the Preaversive Stimulus

All 15 *S*s showed an increase in the B/A ratio between Sessions 3 and 4 and Sessions 9 and 10, as can be seen in Fig. 1. Post hoc analysis of this significant sessions effect, $F\ (3,\ 36) = 6.30,\ p < .01$, reveals that this experimental effect occurred rapidly and did not change significantly between Sessions 5 and 6 and 9 and 10; i.e., the mean B/A ratio for Sessions 5 and 6 is significantly greater than the ratio during Sessions 3 and 4, but does not differ significantly from that in Session Blocks 7 and 8 or 9 and 10 (see Fig. 1). A Newman-Keuls post hoc analysis of the significant Group × Session interaction, $F\ (6,\ 36) = 2.92,\ p < .025$, revealed that this was due to the extreme elevation of the UPL group during Sessions 7 and 8 (see Fig. 1).

To test the criticism that the statistical significance achieved in the previous analysis may be due only to spuriously high ratios on a minority of the trials, the dependent variable was changed to the number of trials per two-session block with a B/A ratio greater than 1.0. A two-factor mixed design (Bruning & Kintz, 1968) revealed a significant sessions effect, F (3, 36)-5.79, $p < .005$.

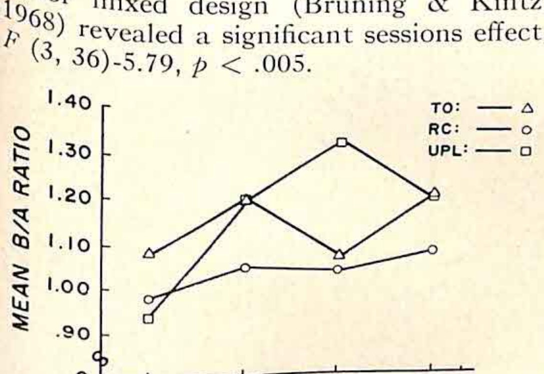


FIG. 1. Mean facilitation ratios B/A (B = number of responses during preaversive stimulus, A = number of responses during 30 sec. preceding preaversive stimulus) as a function of sessions when the aversive event is timeout (TO), response cost (RC), and unavoidable point loss (UPL).

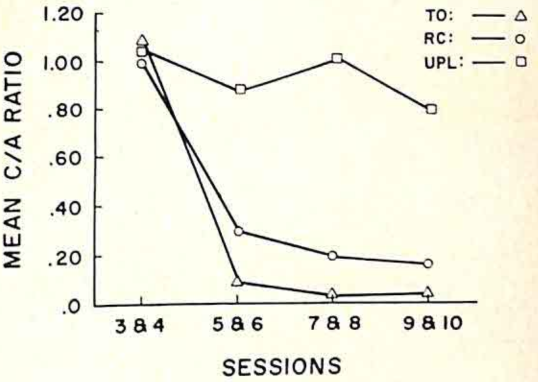


FIG. 2. Mean suppression ratios C/A (C = number of responses during aversive event, A = number of responses during 30 sec. preceding preaversive stimulus) as a function of sessions when the aversive event is timeout (TO), response cost (RC), and unavoidable point loss (UPL).

Effects of the Aversive Stimulus

Figure 2 depicts the mean C/A ratios across sessions for the three experimental groups. Inspection of Fig. 2 reveals rapid and sustained suppression of responding of *S*s during response-cost or timeout periods and little effect due to the presentation of noncontingent point loss. The reliability of these observations is indicated by the significant group effect, $F\ (2,\ 12) = 16.08,\ p < .001$, session effect, $F\ (3,\ 36) = 34.79,\ p < .001$, and Group × Session effect, $F\ (6,\ 36) = 6.14,\ p < .005$.

Effects of the Postaversive Stimulus Condition

No significant difference was noted between *S*s' response rates during the baseline period and those immediately following termination of the aversive stimulus.

DISCUSSION

Previous research has demonstrated response facilitation in the presence of a pretimeout stimulus using animal *S*s (Ferster, 1958; Leitenberg, 1966) and human *S*s (Trenholme et al., 1969). The present experiment replicates this finding and confirms the extension of this relationship to human *S*s.

In addition, this investigation provides evidence for generalizing the facilitation effect during stimuli preceding onset of response cost and unavoidable point-loss periods. Even though all aversive events were not punishing (operationally speaking), all did produce the facilitation effect, making facilitation a phenomenon of surprising generality.

Several theoretical explanations may be advanced to account for this effect. Trenholme et al. (1969) have applied frustration theory to account for a similar result. Their explanation is that frustration, determined by the encounter with a period of nonreinforcement following a period of positive reinforcement, becomes conditioned to the signal preceding timeout. This energizes concurrent operant behavior; hence, results in acceleration of responding. If one adopts this position, he predicts response acceleration immediately following the "frustrating" event, which (in the present investigation) is revealed in a *D*/*A* ratio greater than 1. This latter effect was found not to be significant in the present investigation.

An alternative explanation successfully accounts for the acceleration effect during the preaversive stimulus, and is consistent with the fact that all three groups facilitated about equally. Pliskoff (1961, 1963)—by employing a multiple VI, VI schedule of positive reinforcement—studied the effect on response rate of shifts from one VI schedule of reinforcement to another VI schedule. In this procedure, *S*s were exposed to a VI schedule; then, following a signal, switched to a VI schedule with a different time interval. Pliskoff (1961) reported that

> a) a VI schedule of reinforcement will generate a decreased response rate in a short, discriminated interval before a shift to a higher-frequency-of-reinforcement VI schedule, and b) a VI schedule of reinforcement will generate an increased response rate in a short, discriminated interval before a shift to a lower-frequency-of-reinforcement VI schedule [p. 385].

This finding is applicable directly to the obtained result of response acceleration during a signal preceding timeout, response cost, and point-loss periods. When timeout is employed, *S* proceeds from a higher to lower (zero) frequency-of-reinforcement schedule. The same is true of response cost and unavoidable loss, for (during the cost periods) the "net gain" of reinforcement (positive reinforcement minus point loss) is less than during the no-cost period.

The nonsignificant group effect is understandable because the overall net loss of points per trial for the three groups was the same. On the average, three point deliveries were scheduled during each aversive stimulus presentation; therefore, for the TO group the net loss was 3 points per trial, or 15 points per session. Net loss for the RC group resembled that for the TO group because all but one *S* stopped responding during the response-cost stimulus. Although earning on the average 15 points per session during the aversive stimulus, *S*s in the UPL group lost 15 points per session (3-point loss per trial), for a net loss of 15 points per session. Because all *S*s were subjected to similar relative net losses of points during each session, one would expect, on the basis of Pliskoff's (1961, 1963), data similar acceleration ratios.

The fact that *S*s in the TO and RC groups suppressed response rate during the yellow light and the UPL group did not (see Fig. 2) reveals the importance of the punishment contingency: when point loss is contingent upon an operant response, that response rate decreases; when point loss is noncontingent, there is no noticeable effect.

Analysis of the trials manifesting a *B*/*A* ratio greater than 1.0 was conducted to extend the overall grouped data analysis to individual trials within sessions. Although this analysis replicated the significant session effect, inspection of the data revealed the variability between *S*s and within individual *S*s. While 12 of the 15 *S*s demonstrated an increase in number of trials with *B*/*A* greater than 1.0 between Sessions 3 and 4 and Sessions 9 and 10, inspection of individual *S*s revealed considerable fluctuation from session to session and the fact that only 2 *S*s showed acceleration on all 10 trials during Sessions 9 and 10. The *E* who is interested in predicting the behavior of individuals must consider the problem of generalizing significant results from a general effect to the individual session—finally, to single trials. The degree of confidence one can place in his predictions decreases as he proceeds from general to specific effects.

In the Trenholme et al. (1969) study, *S*s were initially trained to respond on a DRL schedule; that is, initially a lever-pressing response was reinforced only if it followed the previous response by more than 20 sec. Then, *S*s were placed on a VI 1.33-min. schedule, except that no response was reinforced unless at least 5 sec. had elapsed since the last response. Both the initial DRL schedule and the 5-sec. contingency were designed to lower response rates. When a 30-sec. CS was followed by a 5-min. timeout period, increased responding occurred during the CS. The authors noted that—on the base line used—the increased response rates led to increased reinforcement rates.

The present experiment extends the Trenholme et al. (1969) study in several ways. For example, increased reinforcement during the CS is not necessary for facilitation to occur because *S*s could not—in the present experiment—increase their reinforcement rates during the CS.

In addition, response facilitation occurred during the preaversive stimulus in the present experiment in spite of the relatively high base-line rates on the VI 10-sec. schedule. Previous investigators have employed relatively long intervals between reinforcement deliveries, e.g., VI 7 min. (Herrnstein, 1955, in Ferster, 1958, p. 3), VI 3 min. (Ferster, 1958), VI 2.33 min. (Leitenberg, 1966), and VI 1.33 min. (Trenholme et al., 1969). Response patterns of *S*s exposed to a schedule of reinforcement with long mean intervals between reinforcers tend to show moderate-to-low rates of responding. In this experiment all *S*s demonstrated moderate-to-high response rates, and the CS still enhanced the base-line rate in spite of the increased likelihood of interference due to ceiling effects.

REFERENCES

Adelman, H. M., & Maatsch, J. L. Learning and extinction based upon frustration, food reward, and exploratory tendency. *Journal of Experimental Psychology*, 1956, **52**, 311–315.

Baer, D. M. Laboratory control of thumbsucking by withdrawal and representation of reinforcement. *Journal of the Experimental Analysis of Behavior*, 1962, **5**, 525–528.

Baron, A., Kaufman, A., & Stauber, A. Effects of instructions and reinforcement-feedback on human operant behavior maintained by fixed-interval reinforcement. *Journal of the Experimental Analysis of Behavior*, 1969, **12**, 701–712.

Bruning, J. L., & Kintz, B. L. *Computational handbook of statistics*. Glenview, Ill.: Scott, Foresman, 1968.

Catania, A. *Contemporary research in operant behavior.* Glenview, Ill.: Scott, Foresman, 1968.

Estes, W. K., & Skinner, B. F. Some quantitative properties of anxiety. *Journal of Experimental Psychology*, 1941, **29**, 390–400.

Ferster, C. B. Withdrawal of positive reinforcement as punishment. *Science*, 1957, **126**, 509.

Ferster, C. B. Control of behavior in chimpanzees and pigeons by timeout from positive reinforcement. *Psychological Monographs*, 1958, **72**(8, Whole No. 461).

Fleschler, M., & Hoffman, H. S. A progression for generating variable-interval schedules. *Journal of the Experimental Analysis of Behavior*, 1962, **5**, 529–530.

Holz, W. C., Azrin, N. H., & Ayllon, T. Elimination of behavior of mental patients by response-produced extinction. *Journal of the Experimental Analysis of Behavior*, 1963, **6**, 407–412.

Kaufman, A., & Baron, A. Suppression of behavior by timeout punishment when suppression results in loss of positive reinforcement. *Journal of the Experimental Analysis of Behavior*, 1968, **11**, 595–607.

Leitenberg, H. Conditioned acceleration and conditioned suppression in pigeons. *Journal of the Experimental Analysis of Behavior*, 1966, **9**, 205–212.

McMillan, D. E. A comparison of the punishing effects of response-produced shock and response-produced timeout. *Journal of the Experimental Analysis of Behavior*, 1967, **10**, 443–449.

Morse, W. H., & Herrnstein, R. J. The maintenance of avoidance behavior using the removal of a conditioned positive reinforcer as the aversive stimulus. *American Psychologist*, 1956, **11**, 430. (Abstract)

Nigro, M. R. Punishment of an extinguishing shock-avoidance response by timeout from positive reinforcement. *Journal of the Experimental Analysis of Behavior*, 1966, **9**, 53–62.

Pliskoff, S. S. Rate-change effects during a pre-schedule-change stimulus. *Journal of the Experimental Analysis of Behavior*, 1961, **4**, 383–386.

Pliskoff, S. S. Rate change effects with equal potential reinforcements during the "warning stimulus." *Journal of the Experimental Analysis of Behavior*, 1963, **6**, 557–562.

Trenholme, I. A., Baron, A., & Kaufman, A. Effects of signaled timeout from, and loss of monetary reinforcement on, human operant behavior. *Psychonomic Science*, 1969, **15**, 295–296.

Weiner, H. Some effects of response cost upon human operant behavior. *Journal of the Experimental Analysis of Behavior*, 1962, **5**, 201–208.

Weiner, H. Response cost and the aversive control of human operant behavior. *Journal of the Experimental Analysis of Behavior*, 1963, **6**, 415–421.

Weiner, H. Modification of escape responding in humans by increasing the magnitude of an aversive event. *Journal of the Experimental Analysis of Behavior*, 1964, **7**, 277–279. (a)

Weiner, H. Response cost and fixed-ratio performance. *Journal of the Experimental Analysis of Behavior*, 1964, **7**, 79–81. (b)

Weiner, H. Cost to payoff contributions to preference under fixed-ratio contingencies with humans. *Psychological Reports*, 1967, **20**, 31–35.

Weiner. H. Conditioning history and the control of human avoidance and escape responding. *Journal of the Experimental Analysis of Behavior*, 1969, **12**, 1039–1043. (a)

Weiner, H. Controlling human fixed-interval performance. *Journal of the Experimental Analysis of Behavior*, 1969, **12**, 349–373. (b)

Willoughby, R. H. The effect of timeout from positive reinforcement on the operant behavior of children. *Journal of Experimental Child Psychology*, 1969, **7**, 299–313.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received February 2, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 388–396

ACQUISITION AND FORGETTING OF HIERARCHICALLY ORGANIZED INFORMATION IN LONG-TERM MEMORY [1]

THOMAS O. NELSON [2]
University of Washington

EDWARD E. SMITH
Stanford University

Sixty university undergraduates learned a simple hierarchy of connections between letters, and 3 wk. later had a retention test and subsequent relearning. The findings were: (*a*) Both acquisition and retention were better when the hierarchy was learned graphically (i.e., a visual presentation of the entire hierarchy) rather than as a list of associations, although the effect was greater on acquisition than on retention. (*b*) Neither acquisition nor retention was affected by whether the ordering of connections in the hierarchy changed or remained the same across learning trials, suggesting that the internal representation of the hierarchy is not an isomorphic "photographic" image of the physically presented hierarchy. (*c*) Relearning performance and the pattern of retention-test errors suggested the manner in which a hierarchy disintegrates during forgetting from long-term memory.

Both cognitive psychologists (e.g., Bower, 1970) and traditional interference theorists (e.g., Postman, 1971) have stressed the importance of organizational factors during learning. Perhaps due to the analyses of syntax (e.g., Chomsky, 1957) and semantics (e.g., Collins & Quillian, 1969), one obvious kind of organization that has aroused the interest of experimental psychologists is hierarchical organization. Three major studies (Bower, Clark, Lesgold, & Winzenz, 1969; Crothers, 1969; Wortman & Greenberg, 1971) have examined some effects of hierarchical organization on learning, but none of these studies has actually examined the acquisition of unknown hierarchical information which is presented in a graphical hierarchy at the time of study.

Crothers (1969) investigated various ways that a learner could acquire associations from a hierarchy of nonmeaningful CVCs. His method consisted of presenting individual associations from the hierarchy rather than presenting the entire hierarchy at once. Consequently, *S* may never have even noticed the hierarchical properties of the to-be-learned information; if the hierarchical organization became apparent at all, it was only indirectly (via induction) as the course of learning progressed.

Wortman and Greenberg (1971) had the learner acquire a set of meaningful words with hierarchical properties, but they too used a procedure where *S* could not necessarily make direct use of the hierarchical organization (i.e., the entire hierarchy was not presented all at once).

Only the study by Bower et al. (1969) employed the simultaneous mode of presentation, displaying all of the items at the same time in a graphical hierarchy. However, the items in the Bower et al. hierarchy were common words which comprised an a priori hierarchy. That is, words in an order such as MINERAL, STONE, MASONRY, LIMESTONE come from a natural hierarchical organization which the learner knows well in advance of the experimental session—therefore the task is more one of utilizing an already known set of hierarchically organized items rather than one of de novo acquiring a new set of hierarchically organized items.

In summarizing the above research, it becomes obvious that the only study (Bower et al., 1969) which displayed the entire hierarchy at once also utilized to-be-learned hierarchical information that

---

[1] This study was begun while the first author was a National Science Foundation Postdoctoral Fellow at Stanford University, 1970–1971, and was supported, in part, by United States Public Health Service grants to the second author (MH-19705) and first author (MH-21037). We thank Thomas Horan for his throughness in collecting data, and Susan E. Haviland and Colin M. MacLeod for their assistance in analyzing the data.

[2] Requests for reprints should be sent to Thomas O. Nelson, Department of Psychology, University of Washington, Seattle, Washington 98195.

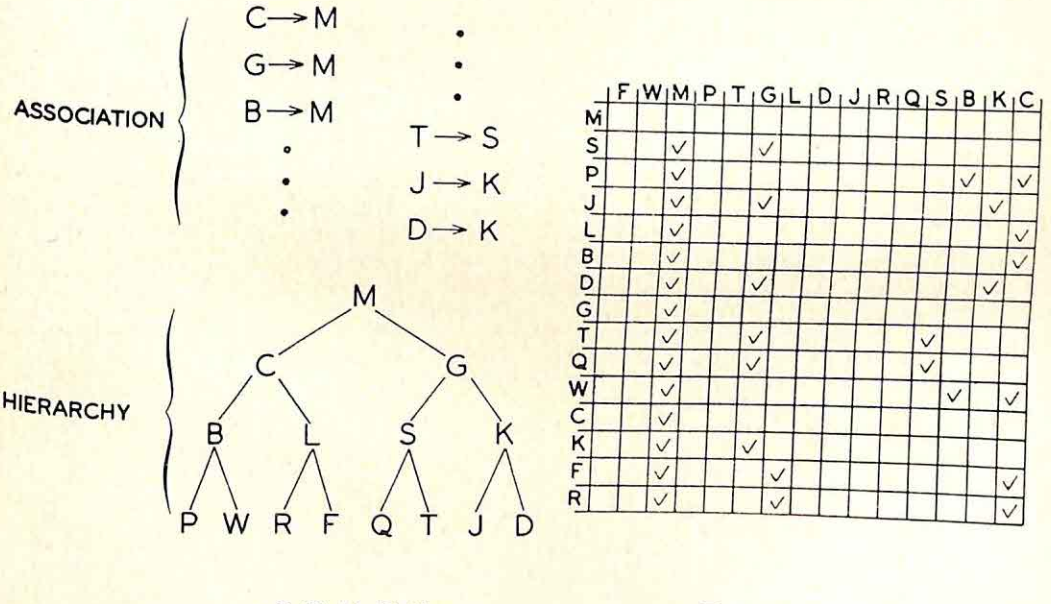
was already known by the learner in advance; the only study (Crothers, 1969) which incorporated unknown to-be-learned hierarchical information did not display the entire hierarchy at once. Thus, one cannot draw any conclusions about the potential advantages of displaying unknown to-be-learned information in a graphically organized hierarchy at the time of study. The present study did employ to-be-learned information which the learner had no knowledge of before entering into the experiment, and the hierarchical structure was made completely obvious by presenting the hierarchy in its entirety during study trials. The hierarchy was simply an array of one-directional connections between letters of the alphabet. In one condition, the connections between letters were shown in a graphical representation of the hierarchy. In a second condition, the learner saw only a list of associations, corresponding to the connections in the hierarchy, without ever seeing the graphical representation of the hierarchy (cf. Fig. 1). Since the actual connections to be learned were identical for both conditions, and were not known in advance, it was possible to determine whether or not learning is affected by a hierarchical representation per se.

Second, and perhaps even more important (cf. Wortman & Greenberg, 1971, p. 239), the present study examined the effects of a hierarchical representation on long-term retention. After attainment of a common learning criterion, *S*s in both the association and hierarchy conditions had a delayed retention test, followed by subsequent relearning. This allowed both (*a*) an assessment of whether a hierarchical representation retards forgetting and (*b*) the determination of how a hierarchy disintegrates during forgetting. It should be noted that forgetting from long-term memory (LTM) and acquisition into LTM are theoretically independent processes; e.g., Olton (1969) found that sentence mnemonics facilitated paired-associate acquisition but had no effect on subsequent forgetting.

METHOD

Design

Acquisition.—There were two crossed between-*S* variables. One variable was the mode of presenta-



	F	W	M	P	T	G	L	D	J	R	Q	S	B	K	C
M															
S			✓			✓									
P			✓										✓		✓
J			✓			✓								✓	
L			✓												✓
B			✓												✓
D			✓			✓								✓	
G			✓												
T			✓			✓						✓			
Q			✓			✓						✓			
W			✓										✓		✓
C			✓												
K			✓			✓									
F			✓				✓								✓
R			✓				✓								✓

FIG. 1. Graphical hierarchy and list of associations used during study trials, and test grid used during test trials (checkmarks indicate correct responses).

tion during study trials. In the Hierarchy condition, *S* saw a graphical hierarchy of one-directional connections between individual letters, such as that shown in Fig. 1. The connections in the hierarchy are determined as follows: Every letter below a given letter is "connected" to that given letter (e.g., the letters S, K, Q, T, J, D are connected to G, whereas only the letters Q and T are connected to S), yielding a total of 34 one-directional connections. In the Association condition, *S* saw a listing (also shown in Fig. 1) of the 34 one-directional connections derived from top to bottom in the hierarchy (cf. "levels down" in Crothers, 1969). A second variable concerned the *ordering* of the connections across study trials. Although the actual identity of the connections remained constant across trials, the ordering of those connections either changed or remained the same, yielding four overall conditions: Hierarchy-Changed, Hierarchy-Same, Association-Changed, and Association-Same. For example, in the Hierarchy-Changed condition, the graphical representation shown in Fig. 1 might be presented on Trial N but during Trial $N + 1$ he parts of the hierarchy having nodal letters C and G might exchange positions, such that during Trial $N + 1$ G (and all letters beneath G) would appear on the left side of the hierarchy while C (and all letters beneath C) would appear on the right side of the hierarchy. Similarly, the ordering of the associations varied across trials in the Association-Changed condition, since the ordering of associations derived directly from the hierarchies in the Hierarchy-Changed condition. The Same versus Changed variable was included to assess the effects of "photographic" organization; i.e., a static visual image of the to-be-learned information could be formed in the Same conditions, whereas the Changed conditions would disrupt the formation of such an image.

Relearning after retention test.—Although the design employed the relearning of both originally learned information and changed information, so that the heading "relearning-transfer" might be more accurate, the term "relearning" is used throughout for the sake of brevity. There were two crossed within-*S* variables which concerned only the lower levels of the hierarchy. Let a *cluster* be defined as a triangular triplet of adjacent letters (e.g., the letters R, L, and F constitute a cluster in the hierarchy shown in Fig. 1). Stated differently, a cluster is comprised of two adjacent connections to the nearest nodal letter (e.g., R → L and F → L). The two relearning variables involved only the four clusters at the bottom of the hierarchy. One relearning variable had the cluster either Changing or remaining the Same as during acquisition. In a Changed Cluster, although the three constituent letters did not vary, the actual connections were Changed (e.g., R → L and F → L from the acquisition phase might change to R → F and L → F during relearning); in a Same Cluster, the connections remained the same as they had been during acquisition. Now let a *superordinate* be defined as the nodal letter immediately above a cluster (e.g., the superordinate of Q, S, T is G while that of R, L, F is C in the hierarchy shown in Fig. 1). The second relearning variable had the superordinate of the bottom-level clusters either Changing or remaining the Same as during acquisition. In the Changed-Superordinate condition, the cluster shifted to a new superordinate (e.g., Q, S, T might move from G to C); in the Same-Superordinate condition, the cluster remained under the same superordinate as in acquisition. Thus, each of the four bottom-level clusters fell into one of four relearning conditions: (*a*) Same-Cluster—Same-Superordinate, (*b*) Same-Cluster—Changed-Superordinate, (*c*) Changed-Cluster—Same-Superordinate, or (*d*) Changed-Cluster—Changed-Superordinate. The Changed condition in both of these relearning variables requires *S* to eliminate, or suppress, the old (acquisition) connections and to form new connections. However, notice that a changed cluster affects only two connections (e.g., R → L and F → L become R → F and L → F) whereas a changed superordinate affects three connections (e.g., Q → G, S → G, and T → G become Q → C, S → C, and T → C).

Subjects

The *S*s were 60 Stanford undergraduates whose participation fulfilled a course requirement. Each of the 60 *S*s was assigned according to order of appearance to one of the four acquisition conditions. These four conditions were rotated so that S_1 was assigned to the first condition, S_2 was assigned to the second condition, and so on.

Procedure

Acquisition.—Each *S* learned 34 one-directional connections between letters. The *S*s in the Association condition were told that the first letter of a given pair (e.g., C → M) "went with" the second letter of that pair; *S*s in the Hierarchy condition were told that every letter below a given letter "went with" that given letter. The study-test method was used and the learning criterion was one errorless test trial. First, *S* had 2 min. to study either the complete hierarchy or the complete list of associations. Since the order of associations was derived from top to bottom in the hierarchy, *S*s in the Hierarchy condition were correspondingly instructed to study the hierarchy from top to bottom. Following study, *S* shadowed digits for 20 sec. to eliminate subsequent recall from short-term memory (STM)—this procedure was used to ensure that in attaining the learning criterion, all of the to-be-remembered information had been stored in LTM rather than in STM (cf. Nelson, 1971b). Then *S* had 4 min. to recall as many connections as he could by making check marks in the appropriate cells of the test grid. The test grid (see Fig. 1) had all of the letters contained in the hierarchy randomly ordered at the edges of the rows and columns (three versions of the test grid were rotated across test trials so that *S* would learn the connections rather than learn the pattern of check marks on the test grid). The *S* made a check mark in the appropriate cell whenever a given row letter "went with" a given column letter (an example of errorless responding is shown in Fig. 1). The *S* was instructed to make exactly 34 check marks, even if he had to guess; approximately 15 sec. before each test trial ended, *E* reminded *S* to be sure that he had made 34 check marks. Following each test trial, *S* had 30 sec. of feedback which consisted of a transparent plastic sheet showing a test grid with a red dot in the center of each of the 34 correct cells; *S* placed the feedback sheet on top of his own test grid to determine where he had made errors. This study-test-feedback sequence continued until *S* had completed one errorless test trial, whereupon the first session terminated. No mention was made of any subsequent retention test.

Prior to learning the main hierarchy, which was composed entirely of consonants, *S* had two complete warm-up trials on a hierarchy comprised of 6 connections between vowels. The presentation during study corresponded to *S*'s condition (e.g., Association-Changed), and the procedure for the warm-up task was the same as that described previously for the main task.

Retention tests and relearning.—The second session took place 3 wk. after the acquisition session. First, *S* had a self-paced free-recall test where he was asked to write down as many of the 34 connections as he could remember, guessing when unsure so that the total number of associations written down was 34. The *S* could use any form of written recall that he wanted (e.g., reconstructing the test grid, listing the associations, or drawing the hierarchy). Following the free-recall retention test, *S* had a self-paced retention test using the test grid. He had to make 34 and only 34 check marks, guessing when unsure. There was no feedback after either retention test. Finally, *S* had one relearning trial on the 34 relearning connections. The *S* was told that some of the connections would be the same as during original learning and some would be different. Only one relearning trial was given because previous research (e.g., Nelson, 1971b) has shown that the magnitude of differential relearning effects is maximal during the first relearning trial and decreases thereafter. The timing and procedure for relearning were the same as described previously for acquisition.

RESULTS AND DISCUSSION

Acquisition

The mean number of acquisition errors for each of the four conditions, over all trials, is shown in the first row of Table 1. Since changing a hierarchy might not be exactly the same as changing a list of associations, suggesting that a two-way analysis of variance may not be entirely appropriate, three orthogonal comparisons were carried out on the data. These three comparisons were: (*a*) Hierarchy-Same and Hierarchy-Changed versus Association-Same and Association-Changed; (*b*) Hierarchy-Same versus Hierarchy-Changed; (*c*) Association-Same versus Association-Changed.

TABLE 1

MEAN NUMBER OF ACQUISITION AND RETENTION ERRORS FOR EACH CONDITION

Stage	Cond.			
	Hierarchy-Same	Hierarchy-Change	Association-Same	Association-Change
Acquisition	11.53	7.60	31.87	34.07
Free-recall retention	6.73	10.00	12.93	14.40
Test-grid retention	4.80	6.87	8.53	9.93

Note.—MS_e for acquisition = 104.75; MS_e for free-recall retention = 60.99; MS_e for test-grid retention = 36.01.

The first comparison showed that the Hierarchy conditions made significantly fewer errors than the Association conditions, F (1, 56) = 79.86, $p < .001$, and this variable is estimated (ω^2) to account for a remarkable 57% of the total variance in acquisition errors. This result strongly supports the notion that a graphically presented hierarchical organization facilitates learning, even when the hierarchical information is not known in advance. There were more than three times as many errors in the Association condition as in the Hierarchy condition, even though the actual connections to be learned were identical for the two conditions.

The second comparison showed that the difference between Same and Changed orders of presentation in the Hierarchy condition was not significant, F (1, 56) = 1.11. The third comparison showed that the difference between Same and Changed orders of presentation in the Association condition was also nonsignificant, $F < 1$. If *S*'s "image" of the to-be-remembered information had been in perfect correspondence with the physical presentation, then changing the presentation order should have produced some disruption during learning because *S* would have had to make repeated alterations. Parenthetically, it can be noted that the Same-Change difference in the Hierarchy condition was not even in the direction predicted by the isomorphic-image notion. Evidently, *S* forms some kind of nonphotographic representation, using the graphical presentation during study trials merely to obtain the information needed to complete his own representation.

An analysis of the number of trials required to reach the acquisition criterion (although these numbers were fairly small) yielded the same conclusions as the abovementioned analysis of errors. The mean numbers of trials to criterion, corresponding to the mean numbers of errors shown in the first row of Table 1, were 1.67, 1.27, 3.07, and 3.13, and the corresponding *F*s for the three orthogonal comparisons were 40.79 ($\omega^2 = 41\%$), 1.22, and .03.

Retention Tests

Free recall.—The mean number of free-recall retention-test errors for each condition is shown in the second row of Table 1. The data were analyzed by the same three orthogonal comparisons as described above for the acquisition data.

The difference between Hierarchy-Same and Hierarchy-Changed conditions was not significant, $F\ (1, 56) = 1.31$, nor was the difference between Association-Same and Association-Changed, $F < 1$. Turning to the major comparison of interest, the Hierarchy conditions made significantly fewer errors than the Association conditions, $F\ (1, 56) = 6.91$, $p < .02$, $\omega^2 = 9\%$. This result suggests that a graphically presented hierarchical organization affects not only acquisition into LTM, but subsequent forgetting from LTM as well. Such a parallel result does not occur for all types of organizational devices (e.g., Olton, 1969). One should also notice that although a graphically presented hierarchy does affect forgetting from LTM, the effect does not seem to be as large as on acquisition (compare the two ω^2s).

To further support the notion that *S*s in the Hierarchy condition were actually retaining a graphical organization, it should be mentioned that of the 30 Hierarchy-condition *S*s, all but 4 drew a hierarchy during the free-recall test; only 1 of the 30 Association-condition *S*s drew a hierarchy, even though the associations were derived from the same hierarchy as originally studied by the Hierarchy-condition *S*s. This finding suggests that if a learner is to derive maximal benefits from a hierarchical organization, all of the items in the hierarchy should be presented simultaneously within the framework of a graphical hierarchical representation.

Test grid.—The mean number of errors on the test-grid retention test for each condition is shown in the third row of Table 1. The results essentially parallel those from the free-recall retention test: The Hierarchy conditions made significantly fewer errors than the Association conditions, $F\ (1, 56) = 4.77$, $p < .05$, $\omega^2 = 6\%$; the difference between Same and Changed orders of acquisition presentation was not significant in either the Hierarchy condition or in the Association condition (both $F\text{s} < 1$).

As the third row of Table 1 also shows, the absolute amounts of forgetting from LTM after the 3-wk. retention interval were notably low—the percentage of errors ranged from $4.8 \div 34 = 14\%$ for the Hierarchy-Same condition to $9.9 \div 34 = 29\%$ for the Association-Changed condition. These low rates of forgetting are probably due, in large part, to the use of a rule such as "every letter goes with M" (as in the example hierarchy shown in Fig. 1), which would produce 14 correct responses; additionally, the low rates of forgetting from LTM are typical of studies (Nelson, 1971a, 1971b) that (*a*) take precautions to eliminate recall from STM during acquisition (thereby ensuring that all of the items are, in fact, stored in LTM at the onset of the retention interval) and (*b*) employ self-paced retention tests so that *S* has enough time to recall long-latency responses.

In some situations, the method of having different groups learn to a common criterion may produce artifactual differences in subsequent retention because of subtle differential overlearning effects (cf. Underwood, 1964). If differential overlearning affected retention in the present study, one would expect that individual *S*s with the fastest learning rates (i.e., the fewest acquisition errors) should show the least forgetting from LTM (i.e., have the fewest retention-test errors). Accordingly, a Pearson product-moment correlation (r) between acquisition errors and retention-test errors was computed for each of the four conditions. None of the four rs was significant ($p > .05$); the mean r (using Fisher's r to Z transformation) was + .01. Thus, there is no evidence that differential overlearning affected retention in the present study.

In addition to examining the amount of forgetting from LTM for hierarchically organized information, an analysis of test-grid retention scores was conducted to determine the manner in which a hierarchy disintegrates during forgetting (i.e., the nature of "chunks" in a hierarchy). Two possibilities were investigated: (*a*) forgetting by clusters and (*b*) forgetting by paths. Each possibility will be considered separately.

Forgetting by clusters would imply that the probability of forgetting one connection in a given cluster (e.g., R → L in the cluster R, L, F in Fig. 1) is dependent upon remembering or forgetting the remaining connection in that cluster (e.g., F → L). If forgetting does not occur by clusters, then the probability of forgetting one connection is independent of remembering or forgetting the remaining connection in that cluster. This independence notion implies that

$$P\,(E_{ij})\cdot P(E_{kj}) = P\,(E_{ij} \cap E_{kj}), \quad [1]$$

where $P\,(E_{ij})$ is the probability of an error on the connection $i \rightarrow j$, $P\,(E_{kj})$ is the probability of an error on the connection $k \rightarrow j$ in the cluster i, j, k, and $P\,(E_{ij} \cap E_{kj})$ is the joint probability of an error on connection $i \rightarrow j$ and an error on connection $k \rightarrow j$. One can think of the left side of Equation 1 as providing the predicted probability of a joint error in a cluster, and this predicted probability of a joint error can be compared with the actual probability of a joint error—the right side of Equation 1—obtained from the data of the test-grid retention test. If the two connections in a cluster are forgotten independently, then one should not be able to reject the null hypothesis that the joint-error probabilities are equal. If, however, forgetting of one connection is dependent on forgetting of the other connection in that cluster, then the observed joint-error probability should be greater than the predicted joint-error probability (cf. transitional error probability in Johnson, 1970). The first row in Table 2 shows the observed and predicted probabilities, separately for each condition but collapsed over *S*s and items, *for all clusters in the lower two levels* of the hierarchy—the data are pooled over these two levels because separate analyses at each level yielded equivalent results. The top level of the hierarchy is omitted because extremely few errors were made on connections to the uppermost letter in the hierarchy. A two-tailed test based on the normal approximation to the binomial (Hayes, 1963, p. 284) showed that, for every condition in the first row of Table 2, the observed probability of a joint error exceeded the corresponding probability predicted by the independence hypothesis ($p < .05$). These results require rejection of the independence hypothesis and suggest that hierarchically organized connections disintegrate by clusters, regardless of whether they are learned in a hierarchical representation or as a list of associations.

TABLE 2

OBSERVED AND PREDICTED JOINT–ERROR PROBABILITIES IN CLUSTERS AND PATHS FOR EACH CONDITION

Hierarchical structure	Cond.			
	Hierarchy-Same	Hierarchy-Change	Association-Same	Association-Change
Cluster	.13(.04)*	.32(.14)*	.39(.23)*	.41(.26)*
Path	.10(.04)*	.08(.08)	.20(.18)	.18(.18)

Note.—Main entry is observed joint-error probability and entry in parentheses is joint-error probability predicted by independence hypothesis.

* Difference between observed and predicted joint-error probabilities is significant at $p < .05$.

The second possibility to be considered is that a hierarchy disintegrates by paths (branches). Let a path be defined as a connected sequence of letters going from the bottom of the hierarchy to the top (e.g., in Fig. 1, the sequence P, B, C, M is a path). The uppermost letter will hereafter be ignored because of too few errors on connections including it, and the general terminology for describing a path (from bottom to top) will be f, g, h which includes the connections $f \rightarrow g$, $g \rightarrow h$, $f \rightarrow h$. If forgetting occurs by paths, then the probability of an error on $f \rightarrow h$, $P\ (E_{fh})$, should increase if a link in the path is broken, i.e., if there is an error on either $f \rightarrow g$ or $g \rightarrow h$, $(E_{fg} \cup E_{gh})$. If forgetting does not occur by paths, so that $P\ (E_{fh})$ and $P\ (E_{fg} \cup E_{gh})$ are independent, then

$$P\ (E_{fh}) \cdot P\ (E_{fg} \cup E_{gh}) = P\ [E_{fh} \cap (E_{fg} \cup E_{gh})], \quad [2]$$

where $P\ [E_{fh} \cap (E_{fg} \cup E_{gh})]$ is the joint probability of an error on $f \rightarrow h$ and an error on either $f \rightarrow g$ or $g \rightarrow h$. Similar to the cluster analysis, the left side of Equation 2 provides the predicted probability of a joint error, which can be compared with the actual probability of a joint error—the right side of Equation 2—obtained from the data of the test-grid retention test. These two probabilities should be the same under the null hypothesis of independence. The second row of Table 2 shows the observed and predicted probabilities, separately for each condition, for all paths in the hierarchy (excluding any connections to the uppermost letter of the hierarchy). For all of the conditions except Hierarchy-Same there is no significant difference between the observed and predicted probabilities. The difference in the Hierarchy-Same condition, although quite small, is statistically significant ($p < .05$). Thus, for all conditions except one, hierarchies do not seem to disintegrate by paths. Unfortunately, the situation for the Hierarchy-Same condition is somewhat muddled.[3]

Since hierarchical information disintegrates more by clusters than by paths *regardless* of the mode of presentation (Hierarchy vs. Association), it appears that the difference in retention between the Hierarchy and Association conditions cannot be explained by differential availability of information within "chunks." Instead, the difference in retention is more likely due to differential availability of higher order information, such as the entire "chunk" itself.

Relearning

Performance was quite high during relearning, very likely because the amount of forgetting from LTM during the retention interval was not great—hence, many of the connections (remembered connections) did not actually have to be "relearned" because they had never been forgotten (cf. Nelson, 1971b). The Hierarchy condition made fewer relearning errors ($\bar{X} = .50$) than the Association condition ($\bar{X} = 4.80$), $F\ (1, 56) = 25.79$, $p < .001$; this difference is of little interest, however, since it is undoubtedly affected by the differential forgetting from LTM during the retention interval for the two conditions (see above). Both the main effect of Same versus Changed order of presentation and the Hierarchy-Association × Same-Changed interaction were nonsignificant (both Fs < 1).

The comparisons of interest focus on the relearning errors for each lower level cluster and its superordinate node. For each cluster-superordinate combination (e.g., P, B, W, C in Fig. 1) there could be errors on five possible connections (e.g., P → B, W → B, B → C, P → C, W → C). A four-way analysis of variance on the relearning data employed the following factors: (*a*) Hierarchy versus Association (an acquisition variable), (*b*) Same versus Changed order (an acquisition variable), (*c*) Same versus Changed clusters, and (*d*) Same versus Changed superordinates. The first two were between-*S* variables and the last two were within-*S* variables. As one would expect from the finding just mentioned, the Hierarchy condition made fewer errors ($\bar{X} = .12$) than the Association condition ($\bar{X} = 1.08$) on cluster-superordinate combinations, $F\ (1,\ 56) = 29.15,\ p < .001$. More relevant to the present discussion, the main comparisons of interest concern the error data for Same versus Changed clusters and Same versus Changed superordinates. These data are shown in Table 3. Significantly more errors were made on Changed clusters than on Same clusters, $F\ (1,\ 56) = 6.23,\ p < .025$. This finding supports the hierarchy-disintegration analysis. Evidently, there is some cohesiveness of the connections within a cluster such that any changes in the cluster produce a decrement in performance, both in terms of forgetting from LTM and also in terms of relearning. The situation is quite different for Same versus Changed superordinates. In essence, changing the superordinate involves breaking the *f*, *g*, *h* path. In the hierarchy-disintegration analysis, the results showed (for all conditions except one) that a broken path due to a forgotten connection does not increase the probability of an error on another connection in that path. The relearning results confirm this notion because changing a superordinate during relearning—and hence, breaking the path—did not produce worse performance than keeping the superordinate the same ($F < 1$). This is particularly interesting because changing a cluster, which necessitates changing only two connections, does produce a significant performance decrement whereas changing a path, which necessitates changing three connections, does not produce a significant performance decrement! Thus, the conclusion about the general importance of paths and clusters for hierarchically organized information seems fairly clear: Both the hierarchy-disintegration analysis and the relearning data suggest that clusters are more critical than paths—changing (or forgetting part of) a cluster produces more of a general performance decrement than does changing (or forgetting part of) a path. Finally, it should be mentioned that all other effects in the analysis were nonsignificant ($p > .10$) except for the triple Hierarchy-Association × Same-Changed superordinate × Same-Changed cluster interaction, $F\ (1,\ 56) = 4.95,\ .025 < p < .05$; this triple interaction seems relatively impotent, both in terms of its magnitude and in terms of its theoretical significance, and none of the double interactions related to the triple interaction were significant ($p > .10$).

---

[3] A parallel significance of paths during relearning for the Hierarchy-Same condition would produce a significant interaction between (*a*) Hierarchy-Association, (*b*) Same-Changed order during acquisition, and (*c*) Same-Changed superordinate during relearning. The interaction was not significant, $F\ (1, 56) = 2.05$, $p > .10$. This lack of parallelism suggests that no firm conclusion should be drawn about the significance of paths for the Hierarchy-Same condition until more research is conducted on this specific issue.

TABLE 3

MEAN NUMBER OF RELEARNING ERRORS FOR EACH CONDITION

Superordinate (Path)	Cluster		
	Same	Changed	$\bar{X}$
Same	.42	.82	.62
Changed	.53	.63	.58
$\bar{X}$	.48	.72	

Note.—MS_e for Superordinate = .58; MS_e for Cluster = .60; MS_e for Superordinate-Cluster interaction = .66; number of possible errors = 5.

CONCLUSIONS

Taken together, the following conclusions summarize the present findings about acquisition and forgetting of hierarchically organized information in LTM.

1. For unknown nonmeaningful connections, presentation of these connections in a graphical hierarchy aids learning *much* more than simply listing these connections, even though the to-be-acquired connections are identical for both modes of presentation.

2. Following attainment of a common acquisition criterion, forgetting from LTM is greater for connections learned in a list than for connections learned in a graphical hierarchy. However, this effect does not seem to be as large on forgetting from LTM as on acquisition into LTM.

3. Direct use of the graphical hierarchy occurs during the retention test, as shown by *S*'s method of responding on a free-recall retention test. However, the internal representation of the hierarchy does not seem to be an isomorphic "photographic" image of the physical hierarchy because Changed versus Same

ordering of the connections had no significant effect on either acquisition or forgetting.

4. When a hierarchy (learned either by a graphical presentation or by a list of connections) is forgotten from LTM, it tends to disintegrate more by clusters than by paths. This conclusion receives support from the pattern of errors on the retention test and from the relearning performance on a new, altered hierarchy. In terms of Miller's "chunking" notion (1956), the "chunk" in hierarchically organized information seems to be a cluster rather than a path.

REFERENCES

Bower, G. H. Organizational factors in memory. *Cognitive Psychology*, 1970, **1**, 18–46.

Bower, G. H., Clark, M. D., Lesgold, A. M., & Winzenz, D. Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 323–343.

Chomsky, N. *Syntactic structures.* The Hague: Mouton, 1957.

Collins, A. M. & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 240–247.

Crothers, E. J. Arbitrary CVC hierarchies learned by paired-component presentations. *Journal of Experimental Psychology*, 1969, **82**, 285–289.

Hayes, W. L. *Statistics for psychologists.* New York: Holt, 1963.

Johnson, N. The role of chunking and organization in the process of recall. In G. H. Bower (Ed.), *The psychology of learning and motivation.* Vol. 4. New York: Academic Press, 1970.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **63**, 81–97.

Nelson, T. O. Recognition and savings in long-term memory: Related or independent? *Proceedings of the 79th Annual American Psychological Association Convention*, 1971, 6, 15–16. (Summary) (a)

Nelson, T. O. Savings and forgetting from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 568–576. (b)

Olton, R. M. The effect of a mnemonic upon the retention of paired-associate verbal material. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 43–48.

Postman, L. Organization and interference. *Psychological Review*, 1971, **78**, 290–302.

Underwood, B. J. Degree of learning and the measurement of forgetting. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 112–129.

Wortman, P. M. & Greenberg, L. D. Coding, recording, and decoding of hierarchical information in long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 234–243.

(Received February 7, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 397–407

TRADE-OFF BIAS AND EFFICIENCY EFFECTS IN SERIAL CHOICE REACTIONS

RICHARD G. SWENSSON[1]

Bell Telephone Laboratories, Holmdel, New Jersey

By comparing the speed-accuracy trade-off relation in two-choice serial tasks, this experiment distinguished between: (*a*) *bias effects*, attributable to changes in *S*'s speed-accuracy trade, and (*b*) *efficiency effects*, that reflected performance drawn from different trade-off functions. One task presented each of the five successive visual stimuli immediately after *S* responded, while the other task delayed the stimulus for 1 sec. When each new stimulus appeared immediately after *S*'s response, a trade-off bias could account for the "repetition effect," i.e., faster, but less accurate, performance for repeated than for alternated responses. Compared to *S*'s performance in his other task, removing the delays lowered efficiency by about the same amount at all levels of speed and accuracy. Although it occurred at all serial positions, this effect seems to require a different explanation for responses at the initial position than for those at later positions. Without delays, posterror responses were exceptional; their high error rates apparently reflected *S*'s persistent tendency to insert extra error-correction responses.

In serial tasks, sequential effects on reaction time (RT) increase in magnitude as the response–stimulus (R–S) delay interval decreases from about 1 sec. to 0 (Bertelson, 1961; Bertelson & Renkin, 1966; Burns, 1971; Rabbitt, 1969). Among correct responses, mean RTs are faster to repeated than to nonrepeated stimuli (Bertelson, 1961) and are slower on trials that follow errors (Rabbitt, 1969). Among errors, repeated responses have faster mean RTs than nonrepeated responses (Hale, 1968).

However, these RT effects are made difficult to interpret by *S*s' ability to trade performance accuracy for increased speed (Fitts, 1966; Hale, 1969; Hick, 1952; Howell & Kreidler, 1964; Leonard & Newman, 1965; Pachella & Pew, 1968; Swensson, 1972). Thus a faster mean RT for responses in one sequentially defined category could reflect either: (*a*) a *trade-off bias*, (e.g., a tendency to respond with less evidence about the stimulus identity) or (*b*) some enhanced *efficiency* in *S*s' perceptual or motor processes. To reject either interpretation, *S*'s performance must somehow be equated for accuracy across response categories. Comparability will not be guaranteed by simply forcing *S* to maintain a low overall error rate, because empirically determined trade-off functions (Hale, 1969; Hick, 1952; Pew, 1969; Swensson, 1972) show that any small difference in error rate leads to progressively larger differences in mean RT as *S*'s overall accuracy increases.

A better way to test these alternative interpretations is to compare *S*'s speed-accuracy trade-off functions between the categories of interest. If such categories partition *S*'s responses without regard to accuracy, a bias to make Category A responses faster (but less accurately) than those in Category B would map out identical trade-off functions for *S*'s A and B responses. Failure to obtain superimposed trade-off functions would imply that A and B responses were not performed with equal efficiency. Parallel trade-off functions, with Category B slower than A at each level of accuracy, would result if B responses required *S* to perform certain extra processing steps that increased his RT without increasing his accuracy (e.g., a shift in attention or in motor-response preparation). However, trade-off functions would differ in slope, *diverging* as *S*'s accuracy increased, if Categories A and B represented occasions on which *S* processed the stimulus information at different rates.

The model for this experiment attributes all sequential effects to biases in *S*'s trade-

---

[1] Requests for reprints should be sent to Richard G. Swensson, Bell Telephone Laboratories, Holmdel, New Jersey 07733.

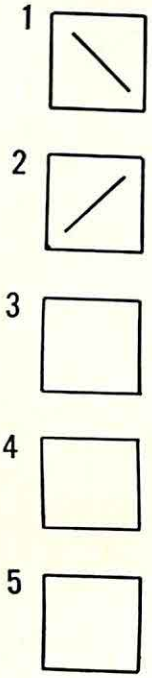


FIG. 1. The visual display. (The stimulus shown in Position 1 required a left-finger response; Position 2 required a right-finger response.)

off or to constant differences in efficiency between response categories. It assumes a serial task with visual choice stimuli displayed in known locations, sufficiently far apart to require a sequence of separate eye fixations. According to the model, *S* must perform three processing steps at each response in the series: he must (*a*) move his eye to the next stimulus location, (*b*) categorize the stimulus and select the appropriate response, and (*c*) execute the motor response he has selected. The model assumes that these steps must be carried out serially, with overlap possible only between execution of one response and the eye movement to the next stimulus. The trade between speed and accuracy is assumed to involve only processes connected with stimulus identification and response selection.

The experiment tested this model in two tasks that required five serial responses to the same visual stimuli, but which differed in their R–S delay interval (0 and 1 sec.). Assuming that the task's R–S delay does not affect the perceptual and decision processes which determine *S*'s responses, the model predicts parallel trade-off functions for all categories of responses in both tasks. If differences in performance efficiency reflect only the amount of *S*'s eye-movement time included in his measured RT, then trade-off functions should be superimposed for responses: (*a*) at the initial serial position in both tasks, and (*b*) at all serial positions when the R–S interval is 1 sec. To insure that *S*'s performances could be matched for accuracy between the two tasks, at each serial position, and between different sequentially defined categories of responses, each *S* performed both tasks under a wide range of trades between speed and accuracy.

METHOD

Task and stimuli—A PDP-7 computer controlled the tasks and presented the stimuli on a cathode-ray tube display.[2] A single trial required five responses, one response to each of the stimuli that appeared in the five square display "windows." The display windows were arranged vertically for one *S* (RG) and horizontally for the other (RS). They were numbered 1 to 5 from top to bottom or left to right, to indicate the sequential order in which the stimuli appeared. The stimulus in each window was a diagonal bar presented in one of the two equiprobable orientations. The *S* responded by pressing a microswitch key under his left or his right index finger, depending on whether the top of the bar tilted 45° to the left or to the right of vertical. Figure 1 shows the vertically arranged display with the "left" stimulus in Window 1 and the "right" stimulus in Window 2.

Except for the stimuli and the square that marked Window 1, all display information remained on the screen continuously. The square marker for Window 1 appeared when *S* pressed a key to begin a new series trial, followed 1 sec. later by Stimulus 1. In the Immediately Serial (IS) task, each response immediately triggered the next stimulus in the next display window. In the Delayed Serial (DS) task, there was a 1-sec. delay between each response and the next stimulus. The computer program ignored any responses that occurred during the R–S delay (impossible by definition in the IS task).

In both tasks, *S*'s fifth response on each series trial automatically produced a feedback display at the top of his screen, which remained until *S* pressed a key. A second keypress then initiated the next series trial. Scoring was automatic; the program assumed one response per stimulus and matched the five responses with the stimuli based on their temporal order of occurrence. The feedback informed *S* of his number of correct responses out of five, the sum of his five RTs, and the sum of his five time costs (for *S* RG, as explained later).

---

[2] Computer programs to perform the experiment and to analyze the data were written by P. E. Collins.

Procedure.—Experimental sessions were 1–2 hr. long and occurred on different days; each was devoted entirely to either the IS or the DS task. Both *S*s, male employees of Bell Telephone Laboratories, Holmdel, N. J., alternated between the IS and DS tasks on successive days. There were 3 practice sessions (2 on the IS task and 1 on the DS task), followed by 8 test sessions. The total numbers of test-session series trials (each of five responses) for the IS and DS tasks, respectively, were about 2,500 and 1,500 for *S* RS and about 3,000 and 2,000 for *S* RG.

Between blocks of 50 or 100 trials within each session, *S* was encouraged to change his performance by changing the relative importance he attributed to speed and accuracy. In the initial block of trials for each session *S*'s primary goal was to achieve high accuracy. Thereafter, *S* was gradually required to speed up his total series RTs (the sum of RTs over his five responses in each series trial) in successive blocks of trials, while obtaining the best accuracy scores consistent with these faster performances. When *S*'s errors had increased to 25%–30% in any single block of trials, the procedure began to require increasingly higher accuracy from *S* in each successive block; he attempted to maintain the fastest total series RTs possible, consistent with these accuracy goals. The *S* RS received no explicit payoffs; his performance goals for each new block of trials were set by instructions concerning error rates or mean total series RT. To change performance goals for *S* RG, explicit monetary payoffs varied between blocks of trials, according to a previously developed procedure (Swensson, 1972).

For each response in the series, *S* RG received a reward of *D* points if correct, nothing if incorrect, and a cost of 1 point for every 10 msec. by which his RT exceeded a "free-delay" interval. As the instructions pointed out, the purpose of this free delay was to remove any incentive for *S* to decrease his RTs by predetermining his responses or guessing (Swensson & Edwards, 1971), since there was no advantage in responding before the free delay elapsed. The free-delay intervals used for the DS (250 msec.) and IS (300 msec.) tasks were estimates of the RT for which *S*'s accuracy would decline to chance level, based on his performance in practice sessions. The value of *D* controlled the relative importance of speed and accuracy for *S*'s payoffs. As *D* decreased in magnitude correct responses became progressively less valuable, so that the time cost gradually began to dominate his net payoffs. Increases in *D* made accuracy more important to the net payoff. The total net payoffs for each block were added to or subtracted from a handicap, calculated from performance in each previous session, to adjust for whatever changes in expected net payoff were produced by the changes in *D*. Points were worth .1¢ each.

RESULTS

Total series performance.—Both *S*s varied the speed and accuracy of their overall series performance across a wide range of trade-offs in each task. The Spearman rank-order correlations over all blocks, between total series RTs and proportion of correct responses, were higher than .943 for both *S*s in each task. However, each serial position contributed only 50 or 100 responses per block. To condense each *S*'s data, so that the speed-accuracy relationship could be examined across various

TABLE 1

ERROR-RATE CRITERIA FOR POOLING BLOCKS OF TRIALS, RESULTING NUMBER OF RESPONSES PER DATA SET, AND MEAN PERFORMANCE IN BOTH TASKS

Measure	Data sets for *S* RG				Data sets for *S* RS			
	1	2	3	4	1	2	3	4
Error-rate criteria								
From	0	.041	.101	.201	0	.051	.101	.251
To	.040	.100	.200	.500	.050	.100	.250	.500
No. responses per serial position[a]								
IS Task	993	945	699	299	699	499	499	599
DS Task	497	549	449	446	200	396	596	289
Error rates								
IS Task	.018	.067	.123	.235	.029	.078	.158	.325
DS Task	.020	.060	.152	.277	.025	.068	.130	.344
Mean RT in msec.								
IS Task	359	348	339	315	439	422	390	325
DS Task	301	293	274	250	383	351	321	264
RT Difference	58	55	65	65	56	71	69	61

[a] The total responses over all serial positions were five times these numbers.

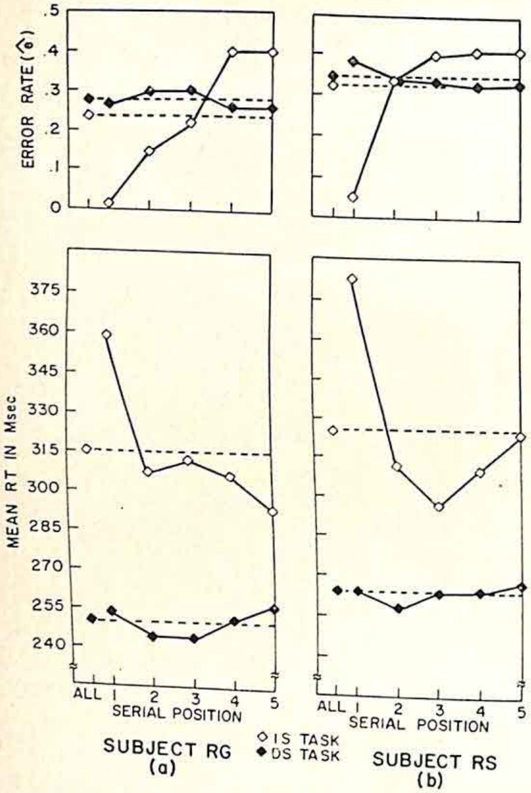


FIG. 2. Mean RTs and error rates at each serial position in the IS and DS tasks from Data Set 4. (Broken horizontal lines indicate mean performance across all serial positions.)

categories of responses, blocks of trials that had similar overall error rates were pooled together for all subsequent analyses. This produced sets of data at four different levels of total series accuracy for each *S*.

Table 1 indicates (*a*) the error-rate criteria that were used for assigning each block of series trials to one of the four data sets, (*b*) the number of responses per serial position that each data set contained in the IS and DS tasks, and (*c*) mean RT and error rates for both *S*s, pooled over all five serial positions in each data set. At comparable levels of accuracy both *S*s' mean RTs were consistently 45–70 msec. faster in the DS task (which used the 1-sec. R–S delay).

Serial position effects.—Initial responses in the IS task were much slower and more accurate than any others. These "initial-position effects" can be seen most dramatically in the data from Set 4, which contained *S*s' fastest and least accurate performances. Figure 2 plots error rate (upper panels) and mean RT over serial positions for Data Set 4. Although RG made 23% errors overall in the IS task, only 1% of his initial responses were errors; responses at Positions 4 and 5 had about 40% errors. The data for RS show a similar pattern. Both *S*s' initial responses in each series also had a mean RT considerably slower than those at later serial positions. However, no such effects emerged in the DS task; Fig. 2 shows that error rates and mean RTs were about the same at each position in the series. These patterns of results repeated for both *S*s in the remaining three data sets, at each level of overall performance accuracy.

The model predicted equally efficient performance for both tasks at the first serial position, because *S* could always fixate this position before the stimulus appeared. However, the model made no predictions about how *S*'s trade between speed and accuracy might change over serial positions or between tasks. Figures 3 and 4 present each *S*'s speed-accuracy trade-offs across the four data sets; they plot a measure of *S*'s accuracy against his mean RT. Accuracy is plotted on a log scale, as the odds favoring a correct response: $(1 - \hat{e})/\hat{e}$, where $\hat{e}$ is the error rate. This measure, the relative likelihood that *S*'s response was correct rather than an error, provides an unbounded scale for accuracy with chance performance (50% errors) at the base line. Corresponding error rates appear on the right-hand ordinate of Fig. 3 and 4.

Connected points in Fig. 3 and 4 (a different shape for each data set) all refer to the same serial position. For each *S*'s DS task (solid points), all five serial positions appear to reproduce the same speed-accuracy trade-off function. However, the display's serial position did seem to affect the trade-off function in the IS task. Numbers beside the open points for the IS task refer to the serial position. The vertical bars around each point bracket plus and minus 1 *SE* of the estimated odds favoring a correct response.

In plots of this type, any data point that

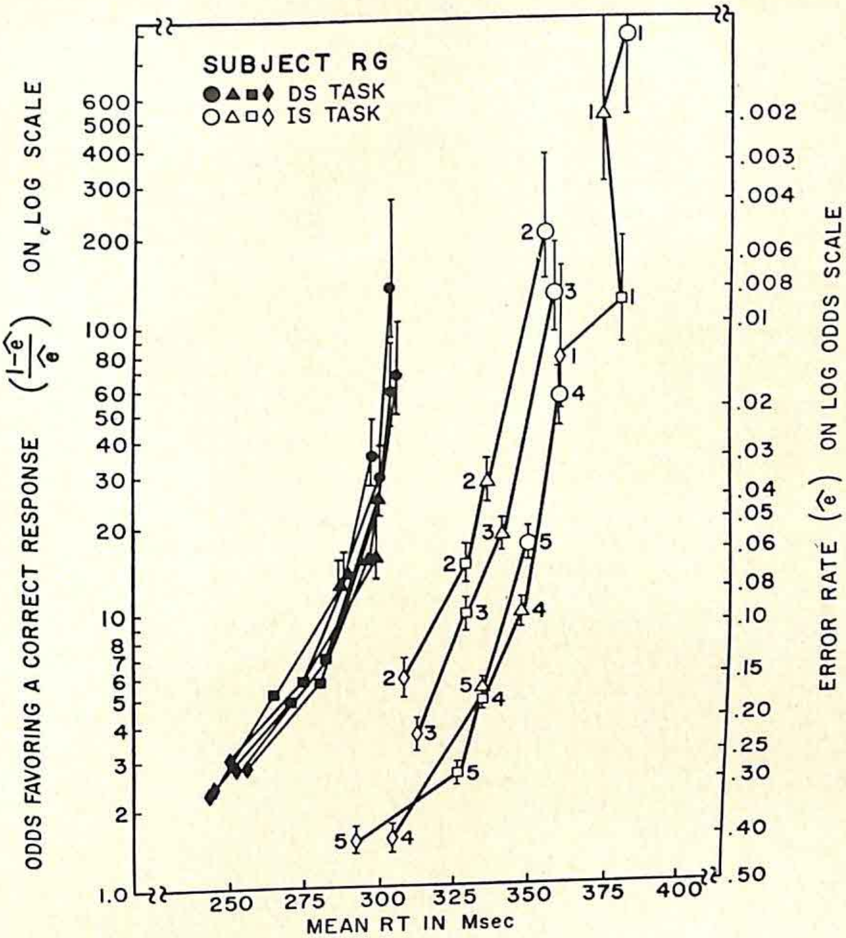


FIG. 3. Speed accuracy trade-offs at each serial position for *S* RG. (The shape of each point indicates the data set to which it refers (Sets 1, 2, 3, 4 = circles, triangles, squares, and diamonds). Open and solid points distinguish performance in the IS and DS tasks. Numbers indicate the serial position represented by each data point for the IS task. Vertical bars around each point bracket plus and minus 1 *SE* of the accuracy-score estimate.)

lies above and/or to the left of some other point indicates more efficient performance. Except for *S*'s initial responses, the curves for each serial position in the IS task roughly parallel those from the DS task, but they are displaced toward slower mean RTs. What appear in Fig. 3 and 4 to be systematic decreases in *S*s' efficiency across Positions 2–5 in the IS task are seen in the next section to be an artifact of the increasing numbers of highly inaccurate posterror responses later in the series. Trade-off functions for *S*s' initial responses in the IS task cannot be located very precisely (because of the extremely low error rates for these responses), but compared to trade-off data from the DS task, these curves appear to be displaced toward slower mean RTs (particularly for RG).

Sequential effects.—A major purpose of the experiment was to test whether sequential effects in these series tasks reflected (*a*) only changes in *S*'s speed-accuracy trade, (*b*) only changes in efficiency, or (*c*) both kinds of effects. Responses at Serial Positions 2–5 in each task were categorized as either *repetitions* (the same as *S*'s response to the previous stimulus) or *alternations* (the opposite of his previous response). A second independent partition

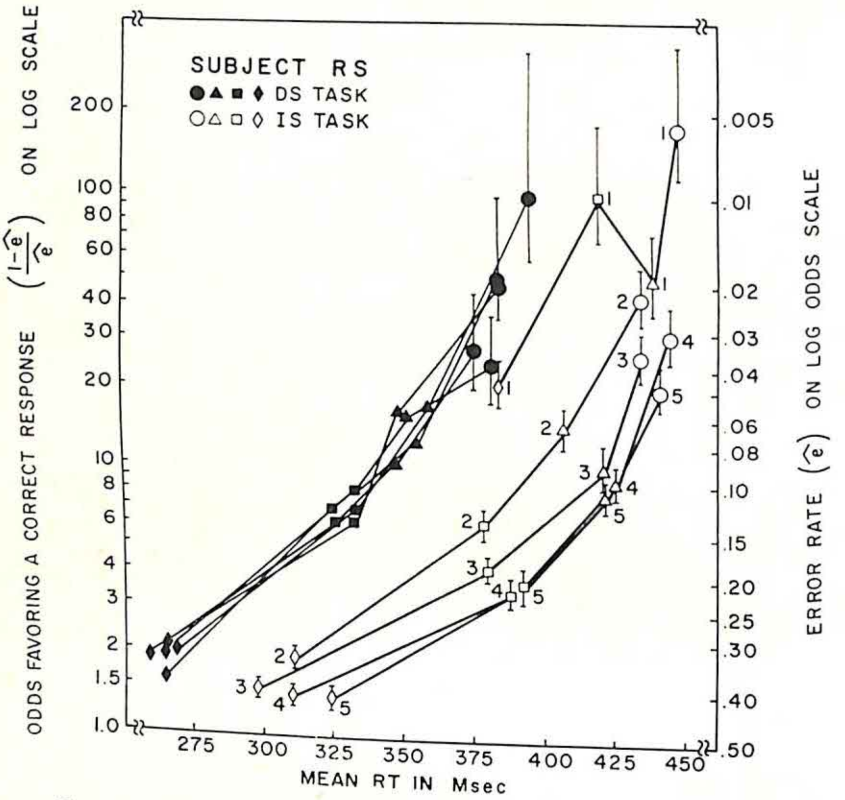


FIG. 4. Speed-accuracy trade-offs at each serial position for *S* RS. (The format is identical with that of Fig. 3.)

further divided each category of responses into two additional subsets: those that *followed correct* responses to the previous stimulus and those that *followed errors*.

Since the two stimuli were equiprobable and independent over serial positions, they directed *S* to make repetition and alternation responses with equal frequency. Any difference in *S*'s speed-accuracy trade between the categories of repetition and alternation responses would produce a response bias. Because *S* would then make one category of responses faster and more frequently than the other, such a trade-off bias would increase the proportion of that response category among his errors—whatever his overall level of performance accuracy. Table 2 presents each *S*'s *repetition*

TABLE 2

REPETITION BIAS FOLLOWING BOTH CORRECT RESPONSES AND ERRORS IN EACH DATA SET FOR THE TWO TASKS

Task	Previous response	*S* RG					*S* RS				
		Data set				All sets pooled	Data set				All sets pooled
		1	2	3	4		1	2	3	4	
IS	Correct	.52	.58	.66	.79	.66	.81	.74	.79	.87	.83
	Error	.29	.18	.17	.08	.14	.25	.21	.10	.29	.24
DS	Correct	.34	.42	.54	.32	.36	.60	.50	.37	.35	.39
	Error	—	.75	.68	.61	.63	—	.67	.72	.50	.55

Note.—The repetion bias is the proportion of the total number of errors in each category that resulted from repetition responses. A value of .5 implies equal numbers of errors contributed by repetition and alternation responses.

bias, the proportion of errors that were contributed by repetition responses, among errors that followed both correct responses and previous errors in each data set for the IS and DS tasks.

Values in Table 2 that are close to 1.0 or 0, respectively, indicate extreme tendencies to repeat or to alternate responses. Biases were small in the DS task, although both *S*s favored alternating responses if they had been correct and repeating the same response if it had been an error. However, in the IS task, *S*s' response biases were considerably more extreme and in the reverse direction. For *S*s RG and RS, respectively, 66% and 83% of the errors that followed correct responses (pooled over all data sets) were contributed by repetitions of their previous responses. Following an error, both *S*s showed a strong tendency to switch responses (to the response they should have made previously); repetition responses constituted only 14% and 24% of the (pooled) errors made by *S*s RG and RS on these trials.

As expected (if they reflected asymmetries in *S*'s trade-offs), the response biases shown in Table 2 were accompanied by systematic differences in performance speed and accuracy between repeated and alternated responses. Whenever there was a repetition bias (i.e., a value in Table 2 substantially larger than .50), then *S*'s repeated responses were faster and less accurate than alternated responses; the reverse was true when *S*'s bias favored alternations. If *S*'s trade-off mechanisms were responsible for all differences between repetition and alternation responses, then the separate trade-off functions for both categories should be superimposed.

Figures 5a and 5b plot the separate trade-off functions obtained from each *S*'s four response categories in the IS task. First consider the trade-off functions following correct previous responses (indicated by the open and filled circles). Although the data points for repetition responses appear shifted toward faster mean RTs and higher error rates (relative to the data points for

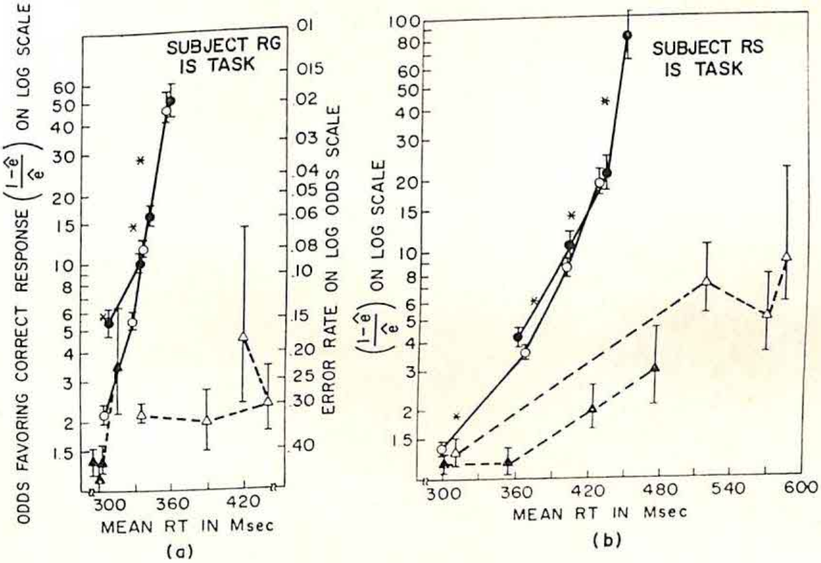


FIG. 5. Speed-accuracy trade-offs from the four types of responses in the IS task for *S* RG (Fig. 5a) and *S* RS (Fig. 5b). (Open and solid points differentiate occasions on which *S* repeated his previous response from those on which he alternated responses. Circles and triangles indicate whether *S*'s previous response had been correct or erroneous. The four points on each curve represent the four data sets. Asterisks indicate performance on all responses in each data set at Serial Position 2. Vertical bars around each point bracket plus and minus 1 *SE* of the accuracy-score estimate.)

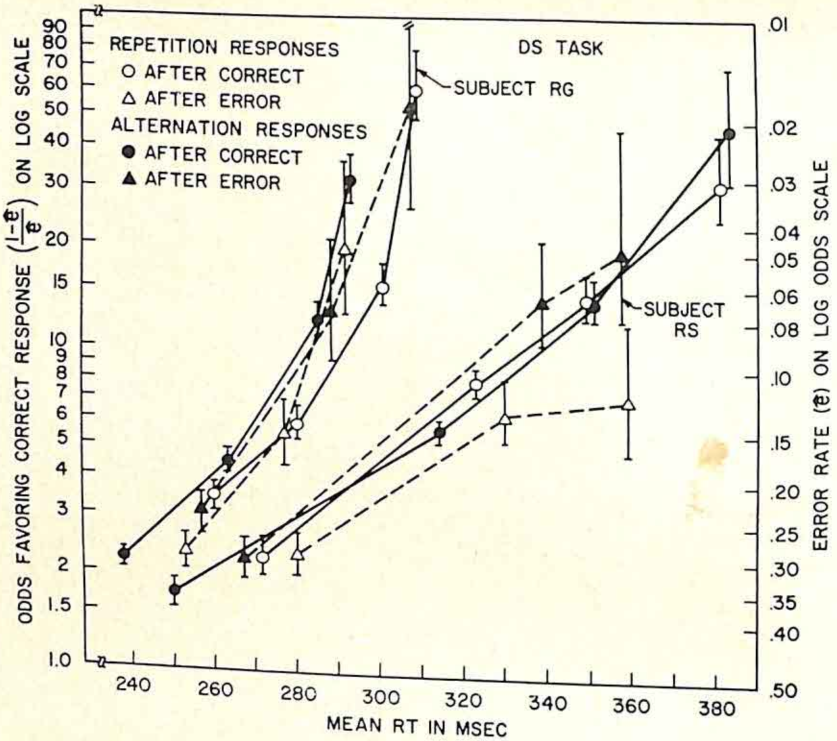


FIG. 6. Speed-accuracy trade-offs for both *S*s' four types of responses in the DS task. (Open and solid points differentiate repeated from alternated responses. Circles and triangles indicate whether *S*'s previous response had been correct or an erroneous. Only three points are plotted for posterror responses (see text). Vertical bars around each point bracket plus and minus 1 *SE* of the accuracy-score estimate.)

alternation responses), both of these functions are generally superimposed in their regions of overlap. Only in RG's fastest and least accurate performance (Fig. 5a) is there any evidence of a systematic difference in these trade-off functions, to suggest that repetition and alternation responses may have differed in efficiency after a correct response.

Both Fig. 5a and 5b clearly indicate that *S*'s performance was degraded after an error in the IS task. Posterror responses had considerably higher error rates in each data set, and mean-RT differences between repetition and alternation responses were also much larger. In all but one case (Set 4 for *S* RS), mean RTs were extremely slow when *S* made an error and then repeated that response, in spite of the high frequency of errors among these responses. Response alternations after an error were usually very fast, particularly for *S* RG. Whether Fig. 5a and 5b suggest a common trade-off function for posterror repetition and alternation responses is questionable. Overlap in the trade-off functions mapped by these two response categories was minimal: only a single data point for RS and no overlap at all for RG.

The degradation in performance that followed an error in the IS task apparently accounts for much of the previously noted decline in performance efficiency from Serial Position 2 to Position 5 (see Fig. 3 and 4). Because error rates increased at the later serial positions, the proportion of posterror responses also increased. When all responses were pooled, this made *S*s' performance appear considerably less efficient at later serial positions. Data points indicated by asterisks in Fig. 5a and 5b represent performance for all responses at Serial Position 2 alone, virtually all of which followed correct responses at the initial position in the series. These data points fall close to the trade-off functions that exclude *S*'s posterror responses.

Figure 6 shows both *S*s' trade-off functions, obtained for the same four response categories in the DS task. Only three points are plotted for posterror responses; because of *S*s' high accuracy such responses were rare in Data Set 1, and none were errors. The trade-off functions were generally similar between all four of *S*s' response categories. Each *S* displayed a small, but consistent, difference in efficiency between repetition and alternation responses; this change in efficiency occurred only after correct responses for RG, and only after errors for RS.

Discussion

Whatever processes underlie *S*s' speed-accuracy trade-off, the proposed model assumed that they would not be affected by the task's R–S delay interval or by the preceding sequence of stimuli and responses. This forces the model to treat all observed performance differences as: (*a*) different trades between speed and accuracy, (*b*) shifts in *S*'s performance efficiency (some constant change in mean RT for all accuracy levels), or (*c*) combinations of both effects. Thus the model predicted parallel trade-off functions for both the IS and DS tasks, for all serial response positions, and for all sequentially defined categories of responses. This prediction was clearly violated for posterror responses in the IS task, which had no R–S delay. As discussed later, the accuracy of these responses increased much more slowly with *S*s' mean RT. However, trade-off functions for all other categories of responses appeared reasonably parallel.

Trade-Off Bias

As previous results (Bertelson, 1961; Bertelson & Renkin, 1966; Rabbitt, 1969) would suggest, the preceding sequence of stimuli and responses had much more pronounced effects when the task included no R–S delay. Then, if *S*'s previous response had been correct, he repeated this response faster and more frequently—but with a higher rate of error—than he alternated responses. This effect apparently represented simply a bias in *S*'s speed-accuracy trade-off, since except for *S* RG's fastest performance (Data Set 4 in Fig. 5a), the separate trade-off functions for repeated and alternated responses were superimposed.

Performance Efficiency

Initial serial position.—At the initial response in each series the IS task was identical to the DS task, except for *S*'s knowledge (presumably reflected in how he organized his performance) that the next stimulus would appear as soon as he responded. Nevertheless, both *S*s made their initial responses much more slowly and accurately in the IS task, and trade-off functions for these responses (seen in Fig. 3 and 4) appear displaced toward slower mean RTs. These data suggest that removing the R–S delay produced some additional processing requirements at the beginning of the IS task, which *forced S*s to spend more time at the first position.

Whatever the precise nature of this initial preparation, *S*s apparently could not complete it during the 1-sec. preparatory delay before the first stimulus appeared. However, the extremely low error rates at this first position demonstrate that *S*s continued processing the stimulus information during at least part of this self-imposed delay.

Later serial positions.—For responses at Positions 2–5 in the series, the IS task shifted *S*s' trade-off functions toward slower mean RTs. Relative to their DS task performance, this shift was 35 to 60 msec. for RG and 45 to 90 msec. for RS. Removing *S*s' inaccurate posterror responses removed the apparent progressive decrease in efficiency from Position 2 to Position 5.

Measured RTs at these positions in the IS task had to include whatever additional time *S*'s eye movements required. Eye movements of predetermined direction and extent require 150 to 200 msec., as estimated by the maximum rate of saccadic fixations (Fitts, 1964) and by eye-movement latency to the onset of a light whose relative position is known (Latour, 1967). Since these estimates are two to four times the change in efficiency observed between the DS and IS tasks, they reject the hypothesis that *S*s had to complete their motor response before they could initiate an eye movement to the next position.

If *S*s began their eye movements and motor responses simultaneously in the IS task, and if motor responses required a mean time of

100 msec,[3] then *S*s would not have completed their eye movements until 50 to 100 msec. after their motor responses (which triggered the stimulus onset). This difference corresponds in magnitude to the observed change in efficiency between the DS and IS tasks.

Small but consistent differences in efficiency appeared between repeated and alternated responses in the DS task, only after correct responses for one *S* and only after errors for the other. Perhaps this task's R–S delay of 1 sec. allowed *S* to make idiosyncratic use of internally generated feedback about whether or not his response was correct—e.g., to prepare one physical response at the expense of the other.

Posterror Responses

In the absence of R–S delays, errors seemed to disrupt *S*s' immediately subsequent performance; although posterror repeated responses were slow and alternated responses were fast, both had much higher error rates than they did after correct responses. These results are consistent with studies of "error correction" responses in serial tasks (Burns, 1971; Rabbitt, 1966; 1967; 1968a; 1968b; 1969). Rabbitt (1966) suggested that *S* may continue to process the stimulus information after he initiates a response. This further processing can trigger a second "more considered" response when it differs from *S*'s first response to the stimulus.

If the task's R–S interval is fairly long, error-correction responses pose no serious problem for the data analysis. If the R–S interval is short, however, there is no unequivocal way to discriminate an error-correction response that occurs after this R–S interval from a fast response made to the next stimulus. The standard procedure (used in the IS task) assumes a single response to each stimulus and records any error correction as an alternated response made to the next stimulus (scored correct with chance-level probability). Thus, the presence of fast error-correction responses generates precisely those effects observed for alternated posterror responses in the IS task, and also found in other experiments with short R–S delays (Rabbitt, 1968b). Error corrections inflate the estimated proportion of response alternations, thereby decreasing mean RT and inflating the error rate for responses in this category.

An *S*'s tendency to correct his errors apparently extends beyond overt error corrections, to all other posterror responses. Posterror responses that fail to correct an error are slower than any other responses (Burns, 1971; Rabbitt, 1966; 1967; 1968b). Rabbitt (1968b) suggested that "*S*s are inhibited from repeating a response which they have just discovered to be wrong [p. 184]." In the present IS task, posterror repetition responses were considerably slower than other responses in all but the fastest performance for one *S*.

In addition to this indirect evidence that error corrections distorted *S*s' posterror performance in the IS task, one phenomenon did clearly indicate that both *S*s did sometimes insert "extra" responses, particularly when they attempted to perform most rapidly. Each *S* occasionally remarked that the feedback display, automatically triggered by the final (fifth) response, occurred too soon—after what he believed was his fourth response. This would happen if *S* made an (unintentional) extra response somewhere in the series, without detecting that he had done so. Unless *S* detects making an error correction and skips the next stimulus, the standard serial RT task records each subsequent response as a response to the stimulus one position ahead in the series.

REFERENCES

BERTELSON, P. Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology*, 1961, **12**, 90–102.

BERTELSON, P., & RENKIN, A. Reaction times to new versus repeated signals in a serial task as a function of response-signal time interval. *Acta Psychologica*, 1966, **25**, 132–136.

BURNS, J. T. Error-induced inhibition in a serial reaction time task. *Journal of Experimental Psychology*, 1971, **90**, 141–148.

FITTS, P. M. Engineering psychology and equipment design. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

FITTS, P. M. Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964.

FITTS, P. M. Cognitive aspects of information processing: III. Set for speed versus accuracy. *Journal of Experimental Psychology*, 1966, **77**, 849–857.

---

[3] Motor response time must include delays of about 25 to 50 msec. for neural conduction (Ruch, 1951), about 40 msec. for muscular contraction (Fitts, 1951), and the push button's travel time. An estimate of about 100 msec. is consistent with these times and also agrees with the minimum duration between discrete repetitive movements, such as finger tapping.

HALE, D. J. The relation of correct and error responses in a serial choice reaction task. *Psychonomic Science*, 1968, **13**, 299–300.

HALE, D. J. Speed-error tradeoff in a three-choice serial reaction task. *Journal of Experimental Psychology*, 1969, **81**, 428–435.

HICK, W. E. On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 1952, **4**, 11–26.

HOWELL, W. C., & KREIDLER, D. L. Instructional sets and subjective criterion levels in a complex information processing task. *Journal of Experimental Psychology*, 1964, **68**, 612–614.

LATOUR, P. L. Internal clocks in the human operator. *Acta Psychologica*, 1967, **27**, 341–348.

LEONARD, J. A., & NEWMAN, R. G. High speed and high accuracy in a keyboard task. *Ergonomics*, 1965, **8**, 281–304.

PACHELLA, R. C., & PEW, R. W. Speed-accuracy tradeoff in reaction times: Effect of discrete criterion times. *Journal of Experimental Psychology*, 1968, **76**, 19–24.

PEW, R. W. The speed-accuracy operating characteristic. *Acta Psychologica*, 1969, **30**, 16–26.

RABBITT, P. M. A. Errors and error correction in choice-response tasks. *Journal of Experimental Psychology*, 1966, **71**, 264–272.

RABBITT, P. M. A. Time to detect errors as a function of factors affecting choice-response time. *Acta Psychologica*, 1967, **27**, 131–142.

RABBITT, P. M. A. Repetition effects and signal classification strategies in serial choice-response tasks. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 232–240. (a)

RABBITT, P. M. A. Three kinds of error-signaling responses in a serial choice task. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 179–188. (b)

RABBITT, P. M. A. Psychological refractory delay and response–stimulus interval duration in serial choice-response tasks. *Acta Psychologica*, 1969, **30**, 195–219.

RUCH, T. C. Motor systems. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

SWENSSON, R. G. The elusive tradeoff: Speed versus accuracy in visual discrimination tasks. *Perception & Psychophysics*, 1972, **12**, 16–32.

SWENSSON, R. G., & EDWARDS, W. Response strategies in a two-choice reaction task with a continuous cost for time, *Journal of Experimental Psychology*, 1971, **88**, 67–81.

(Received February 18, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 408–411

CONDITIONED INHIBITION OF THE RABBIT'S NICTITATING MEMBRANE RESPONSE[1]

HORACE G. MARCHANT, III, FREDERICK W. MIS,
AND JOHN W. MOORE[2]
University of Massachusetts, Amherst

Albino rabbits received Pavlovian conditioned inhibition training of their nictitating membrane response to a light (CS +) and the same light plus a pure tone (CS −). The results of summation and retardation tests, in comparison with pseudo-conditioned inhibition controls, indicated that the tone acquired inhibitory properties as a function of conditioned inhibition training.

In Pavlov's (1927) conditioned inhibition paradigm, a conditioned response (CR) is established to two separate and distinct stimuli, CS_1 and CS_2. This is followed by a special kind of differential conditioning in which CS_1 continues to be reinforced while a compound consisting of CS_1 and a third stimulus, CS_3, is not. The *S* gradually forms the appropriate discrimination between CS_1 and the compound. One test of the acquired inhibitory properties of CS_3 (summation) consists of presenting it together with CS_2 and assessing CR suppression relative to suitable controls. Presenting CS_3 with CS_2 is essential if one is to conclude that CS_3 has acquired inhibitory properties of its own and not merely as part of a compound or gestalt with CS_1. Another test of the inhibitory properties of CS_3 (retardation) consists of determining whether subsequent excitatory conditioning to this stimulus is impeded as a consequence of conditioned inhibition training.

Rescorla (1969) has detailed criteria which must be met before it can be asserted that a given training procedure (e.g., simple differential conditioning, conditioned inhibition training) results in an inhibitory CS. First, it must be demonstrated that systematic nonreinforcement produces a CS with properties opposite in direction to those produced by systematic reinforcement. Thus, following conditioned inhibition training, the addition of the reinforced CS (CS_1) to CS_2 should produce an increase in the likelihood of a CR in a summation test compared with CS_2 alone. The addition of CS_3 to CS_2 should have the opposite effect of lowering the likelihood of a CR relative to CS_2 alone. Rescorla points out that the latter result, while necessary, is not sufficient for establishing an inhibitory role of CS_3. It is essential in addition to rule out the possibility that the suppressive effect of CS_3 in combination with CS_2 is not merely a consequence of generalization decrement, attention shift, or some other factor related to the novelty of the test compound. As one step toward satisfying this criterion, the present experiment included control animals which received the same treatment as conditioned inhibition animals except the UCS occurrences were uncorrelated with trial type during the training phase (pseudo-conditioned inhibition). A retardation test in reacquisition provided an additional check against the attention-shift factor: Following Rescorla's argument, a CS_3 which is sufficiently salient to distract *S*'s attention from CS_2 in the summation test, thereby producing a suppression of the CR to the compound, should reacquire its initial excitatory properties quite rapidly. On the other hand, if CS_3 does possess inhibitory properties, its reacquisition should be retarded.

Since there have been very few English-language reports of studies which have actually employed Pavlov's (1927) conditioned inhibition paradigm to establish a conditioned inhibitor (cf. Rescorla, 1969),

---

[1] This research was supported by National Science Foundation Grant GB 24557.

[2] Requests for reprints should be sent to John W. Moore, Middlesex House, Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01002.

it seemed desirable to subject conditioned inhibition in the rabbit eye-blink preparation to the stringent analysis dictated by the criteria set down in the preceding discussion.

METHOD

Subjects and apparatus.—The *S*s were 27 naive albino rabbits weighing approximately 2 kg. A detailed description of the apparatus is available elsewhere (e.g., O'Malley, Hupka & Moore, 1969). Briefly, up to four animals were run concurrently in individual ventilated and sound-proofed file drawers. Each rabbit was restrained in Plexiglas boxes like those described by Gormezano (1966). A potentiometer mounted on *S*'s head was connected by a small metal hook and silk thread to a nylon loop sutured into the nictitating membrane of the right eye. Movement of the membrane resulted in a dc signal which was amplified and recorded on a Grass 5D oscillograph. The UCS was an ac shock of 2 ma. and 50-msec. duration delivered through two stainless-steel wound clip (Clay Adams, 9-mm.) electrodes attached to the skin of the infraorbital region of the right eye. No eyelid retractors of any kind were employed. Conditioned stimuli were as follows: (*a*) CS_1 was either the onset or offset (for .5 sec.) of two 4.5-v. incandescent lights mounted behind two translucent white plastic screens located side by side directly in front of the animal; (*b*) CS_2 was a .5-sec. burst of white noise at 86 db. SPL; (*c*) CS_3 was a .5-sec. tone of 1,200 Hz. and 84 db. delivered over a speaker located directly in front of *S*. On reinforced trials, the UCS overlapped the final 50 msec. of the CS so that the CS–UCS interval was .5 sec. Conditioned responses were defined as 1-mm. deflections of the oscillograph pen in the direction indicating extension of the nictitating membrane of less than 1 mm. and occurring within the CS–UCS interval.

Design and initial conditioning.—A total of 20 rabbits received conditioned inhibition training, and 7 served as pseudo-conditioned inhibition controls. For 12 of the former and 3 of the latter, CS_1 consisted of light offset rather than light onset. Subsequent analysis indicated that this variation was not related to the outcome of the experiment. In the present experiment, CS_3 as well as CS_1 and CS_2 were reinforced in an initial acquisition phase. The purpose of this variation of Pavlov's (1927) procedure was to reduce the influence of unconditioned or latent inhibitory properties of CS_3 (Lubow & Moore, 1959). Consequently, retardation was assessed in terms of reacquisition.

All rabbits were habituated to the apparatus for a period of 15 min. Conditioning sessions of 99 or 100 trials (depending upon whether two or three CSs were being conditioned on that session) began on the next day, with each rabbit receiving reinforced pairings of each CS (presented separately) with the UCS in an unsystematic order. The intertrial interval in this and all subsequent phases of the experiment was a constant 40 sec. Daily conditioning sessions continued until each animal attained a level of at least 80% level within this time, but the criterion was not rigidly adhered to in the case of this stimulus in view of the fact that it would continue to be reinforced in the next stage of training. The mean number of initial conditioning sessions was 4.3 for conditioned inhibition *S*s and 4.4 for control animals. The experimental design is summarized in Table 1.

Conditioned inhibition training.—Conditioned inhibition training for the experimental *S*s consisted of seven daily sessions of 100 trials: 50 reinforced trials to CS alone and 50 nonreinforced presentations of the compound CS_1 plus CS_3. The two trial types were unsystematically alternated. The control *S*s received the same number and sequence of CS presentations, but each trial type was reinforced only half of the time according to a random schedule.

Summation testing.—On the day following the final day of conditioned inhibition training, all *S*s were given 17 nonreinforced presentations of each of the following six CS combinations: (*a*) CS_3 alone, (*b*) CS_1 alone, (*c*) CS_2 alone, (*d*) CS_1 plus CS_3, (*e*) CS_2 plus CS, and (*f*) CS_1 plus CS_2. These were presented in an unsystematic order, for a total of 102 trials per day for 3 successive days.

Reacquisition.—Following summation testing, all rabbits were given daily conditioning sessions (99 trials/day) to each of the three CSs (33 trials) until the 80% CR criterion to CS_3 was attained.

TABLE 1

SUMMARY OF EXPERIMENTAL DESIGN AND PROCEDURE

Group	Phase 1: 3–6 days	Phase 2: 7 days	Phase 3: 3 days	Phase 4: 1–3 days
Conditioned inhibition	CS_1+ CS_2+ CS_3+	CS_1+ $(CS_1\ \&\ CS_3)-$	$CS_1-\ (CS_1\ \&\ CS_3)-$ $CS_2-\ (CS_2\ \&\ CS_3)-$ $CS_3-\ (CS_1\ \&\ CS_2)-$	CS_1+ CS_2+ CS_3+
Control	CS_1+ CS_2+ CS_3+	$CS_1\pm$ $(CS_1\ \&\ CS_2)\pm$	$CS_1-\ (CS_1\ \&\ CS_3)-$ $CS_2-\ (CS_2\ \&\ CS_3)-$ $CS_3-\ (CS_1\ \&\ CS_2)-$	CS_1+ CS_2+ CS_3+

Note.—Light = CS_1; noise = CS_2; tone = CS_3. Reinforced stimuli are denoted by plus (+) signs; nonreinforced stimuli are denoted by minus (−) signs. For the conditioned inhibition groups, $n = 20$; for the control group, $n = 7$.

TABLE 2

MEAN PERCENTAGE OF CONDITIONED RESPONSES (CRs) TO EACH TEST STIMULUS DURING SUMMATION TESTING

Group	Test stimulus					
	CS_1	CS_1 plus CS_3	CS_3	CS_1 plus CS_2	CS_2	CS_2 plus CS_3
Conditioned inhibition	58	40	18	59	43	31
Control	34	57	34	62	44	44

RESULTS AND DISCUSSION

Differential conditioning.—As a consequence of pretraining, mean levels of conditioned responding to CS_1 and the CS_1 plus CS_3 compound were initially high: 84% and 90% CRs on the first day of conditioned inhibition training, respectively. Responding to CS_1 remained high, averaging 87% CRs over the last 3 days, whereas responding to the compound decreased progressively to an average of 63% CRs for the last 3 days of conditioned inhibition training. By contrast, the corresponding mean CR rates for the control group averaged over the last 3 days of pseudo-conditioned inhibition training were 84% to CS_1 alone and 97% to CS_1 plus CS_3. A second-order difference t test indicated that the differential in CR frequency between CS_1 and CS_1 plus CS_3, pooled over the last 3 days, was significantly different for conditioned inhibition and control animals, $t(25) = 3.18$, $p < .01$. It is therefore evident that the compound CS acquired response-suppressive properties for conditioned inhibition *S*s as a consequence of systematic nonreinforcement.

Summation testing.—Table 2 summarizes the results of summation testing. Each entry is the mean percentage of CRs to each test stimulus pooled over 3 test days. Considering the results for conditioned inhibition animals, the important contrasts are between CS_1 plus CS_2 and CS_2 alone and between CS_3 plus CS_2 and CS_2 alone. As a conditioned excitator, CS_1 was expected to increase the likelihood of a CR when combined with CS_2 in comparison with CS_2 alone. This result was obtained and was significant, sign test, $p < .002$. Conversely, responding to the combination of CS_1 plus CS_2 should have exceeded responding to CS_1 alone as well, but this was not the case. The failure to observe excitatory summation with respect to CS_1 was probably due to the comparatively high level of responding to CS_1 alone (58%), and any further increment in CR strength by the addition of CS_2 would have been difficult to achieve. Hull's (1943, p. 178) principle of physiological addition anticipated this problem by postulating a law of diminishing returns for stimulus summation. By contrast, excitatory summation of CS_1 plus CS_2 with respect to both CS_1 and CS_2 alone appears to have held for control *S*s. In this case, responding to CS_1 alone was low enough (34%) for the summation effect to emerge, sign test, $p = .02$. Given these arguments, it appears reasonable to conclude that the observed difference in responding to CS_1 plus CS_2 and CS_2 alone by conditioned inhibition animals was a valid instance of excitatory summation.

The acquired inhibitory properties of CS_3 for conditioned inhibition animals is indicated by the decrease in responding to CS_2 produced by the addition of CS_3, sign test, $p < .002$. The validity of this inhibitory summation effect is supported by a second-order difference t test contrasting the differential in responding between CS_2 and CS_2 plus CS_3 of the conditioned inhibition group with that of the control group, $t(25) = 3.50$, $p < .01$. This comparison provides the most compelling evidence of inhibitory summation because the level of responding to CS_2 in the two groups was practically identical (43% and 44% for conditioned inhibition and control groups, respectively). For this reason, it is clear that the significant second-order difference was the result of greater response suppression produced by the addition of CS_3 to CS_2 in the conditioned inhibition group. Had the base lines in the two differed appreciably, differences between responding to CS_2 and CS_2 plus CS_3 in the two groups might have proved difficult to interpret. The finding that the addition of CS_3 to CS_2 resulted in suppression, whereas the addition of CS_1 to CS_2 produced enhancement, satisfies the most important criterion of conditioned inhibition proposed by Rescorla (1969).

The inhibitory function of CS_3 for conditioned inhibition is also supported by the reduced level of responding to CS_1 plus CS_3 compared with CS_1 alone, sign test, $p < .03$. This result indicates that the differential in conditioned responding to the two training CSs carried over into the extinction test phase.

The greater level of responding to the combination of CS_1 plus CS_3 over single CSs in the control group, while not statistically significant, suggests that CS_3 acquired an excitatory function as a result of pseudo-conditioned inhibition training in which this stimulus was reinforced on half of the trials.

Reacquisition.—Having demonstrated an inhibitory summation effect for the presumed conditioned inhibitor CS_3, the primary purpose of a retardation test is to assess the validity of the attention-shift interpretation of summation presented in the introduction. The mean percentages of CRs on the first day of reacquisition to CS_3, CS_1, and CS_2 were 75, 90, and 96 for conditioned inhibition animals and 89, 83, and 93 for control *S*s, suggesting retardation. Retardation of reacquisition was assessed by a second-order difference test in which the differential in responding between CS_2 and CS_3 for conditioned inhibition rabbits was contrasted with the control group, $t\ (25) = 2.05$, $p < .05$. The mean percentage of CRs to CS_2 and CS_3 on the last extinction test day were 32% and 17% for conditioned inhibition rabbits and 30% and 21% for controls. Since the two groups did not differ appreciably at this stage, the observed retardation of reacquisition of CS_3 relative to CS_2 would appear to be a genuine instance of retardation and not merely carry over from extinction.

A second-order difference test contrasting responding to CS_3 in relation to CS_1 for conditioned inhibition and control groups also indicated significant retardation, $t\ (25) = 2.31$, $p < .05$. However, this effect could have been due, at least partially, to a carry-over from the last day of extinction in which the conditioned inhibition group gave 58% CRs to CS_1 against 28% for the control group.

CONCLUSION

Differential conditioning in Pavlov's (1927) conditioned inhibition paradigm resulted in a CS which met all criteria of inhibition outlined by Rescorla (1969): (*a*) Its effect was opposite in direction to a conditioned excitator in the summation test. (*b*) Comparison with pseudo-conditioned inhibition controls and the observed retardation of reacquisition essentially eliminates an interpretation of the summation results in terms of "attention-shift" or other factors not related to conditioned inhibition training.

REFERENCES

Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

Gormezano, I. Classical conditioning. In J. B. Sidowski (Ed.), *Experimental methods and instrumentation in psychology.* New York: McGraw-Hill, 1966.

Lubow, R. E., & Moore, A. U. Latent inhibition: The effect of nonreinforced preexposure to the conditioned stimulus. *Journal of Comparative and and Physiological Psychology*, 1959, **52**, 415–419.

O'Malley, P., Hupka, R. B., & Moore, J. W. Auditory differential conditioning of the rabbit nictitating membrane response: I. Effects of mixed- and separate-phase training. *Psychonomic Science*, 1969, **15**, 305–306.

Pavlov, I. P. *Conditioned reflexes.* London: Oxford University Press, 1927.

Rescorla, R. A. Pavlovian conditioned inhibition. *Psychological Bulletin*, 1969, **72**, 77–94.

(Received February 22, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 412–418

PROACTIVE INTERFERENCE IN SHORT-TERM RECOGNITION AND RECALL MEMORY[1]

WILLIAM M. PETRUSIC[2] AND RICHARD F. DILLON
Carleton University

The rates of proactive interference (PI) buildup under recall and two recognition memory test conditions were compared over four Brown-Peterson trials for 135 *S*s. Whether the distractors were previous targets or new words from the same encoding class as the targets did not influence recognition performance. Although uncorrected performance under both recognition conditions was better than under recall, there was significant PI in all conditions, and the rates of PI buildup over trials were the same. When *S*s were shifted to recognition on Trial 5 after recall on Trials 1–4, performance was comparable to *S*s tested for recognition on all five trials. For recall on Trial 5, performance was best for the group maintained on recall, lower for *S*s who had seen old distractors, and lowest for *S*s who had seen new distractors.

The accumulation of proactive interference (PI) over the first few trials in the Brown-Peterson paradigm is well established (e.g., Keppel & Underwood, 1962; Murdock, 1961; Wickens, 1970). However, it is not known whether PI is present in the Brown-Peterson situation when memory is tested using a recognition measure. Hawkins, Pardo, and Cox (1972) tested recognition on Trial 2 following recall on Trial 1. While Hawkins et al. demonstrated that recognition following recall was poorer than recognition on the first trial, they did not demonstrate that PI develops while using a recognition test procedure. Thus, the first purpose of this study was to compare the rates of increase in PI over the first few trials under recall and recognition memory test conditions.

A second purpose was to determine the effect of two types of distractors on short-term recognition. Two recognition procedures which required *S*s to select three target items from among three distractors, all from the same encoding class, were used. Distractors for one recognition procedure were targets from the immediately preceding trial, while distractors for the second recognition procedure were previously unused words. From what is known about PI using a recall measure and from the Hawkins et al. (1972) results, little PI should occur in the condition with new distractors, more PI in the condition with old distractors, and the most PI should occur with a recall measure.

The third purpose of the experiment was to test memory using the three memory test procedures after PI had reached a stable level under each test procedure. By allowing PI to achieve a steady state and then either shifting to another memory test procedure or maintaining the same test procedure, it should be possible to examine the relative independence of PI from previous test history. For example, while memory might be good using a recognition procedure and there might be no evidence of PI, what would the level of memory be if a recall test procedure were introduced? Would PI develop during recognition trials, but manifest itself only with a recall test? On the other hand, would PI, accumulated during a series of recall trials, dissipate if a recognition test is introduced? This manipulation is, in many respects, similar to a PI release situation (e.g., Wickens, 1970) with release (or failure to release) due not to a change in recall material, but rather to a change in the type of memory test condition.

---

[1] The research reported here was supported by grants from the Medical Research Council of Canada (MA-4093) and the National Research Council of Canada (APA-7746 and APA-0372). We thank P. D. McCormick for many contributions and L. Little for her assistance. The generous cooperation of the Board of Directors of the Central Canada Exhibition Association is gratefully acknowledged.

[2] Requests for reprints should be sent to William M. Petrusic, Department of Psychology, Carleton University, Ottawa, Ontario, Canada.

METHOD

Subjects.—One hundred thirty-five persons attending the Central Canada Exhibition in Ottawa, Ontario volunteered without pay for the experiment. They were attracted by a sign which asked if they could "do two things at once," and which requested that they "help with our research on thinking." The data from 8 *S*s were discarded because of errors in procedure, and additional *S*s were run. There were 66 males and 69 females, with a mean age of 24.2 yr. ($s = 9.8$ yr.), ranging from 17, the minimum accepted, to 71 yr. Sixty-three *S*s were high school or university students, 11 were housewives, and the remainder were distributed among 16 other occupations.

Recall material.—The *S*s were required to remember three visually presented words on each of five trials. Thirty words taken from the Heise (1965) norms served as recall items for the experimental trials. These words were scaled as moderate to extreme negative on the evaluative dimension and as near to zero as possible on the other two Osgood dimensions. The triads formed from these words are presented in Table 1.

Distractor-task material.—While trying to remember the three words, *S*s were required to perform a number-classifying task. Randomly selected one- and two-digit numbers arranged in 4 Columns × 5 Rows served as distractor material. When one of the distractor slides appeared, *S* read the upper-left number out loud, classified it as high (50 or above) or low (below 50) and as odd or even. The *S* then worked down the left column and continued to classify numbers down the columns from left to right. Instructions emphasized speed and accuracy. No *S* was able to complete all of the numbers during the 16 sec. allotted for each trial. The 17 digit slides for 17 trials appeared in the same order for all *S*s.

Memory test conditions.—Upon completion of the distractor task, *S*s were tested for retention of the three words using one of three procedures. In the ordered recall procedure, *S* had to retrieve the three recall items and write them in the correct order. This procedure was followed when three question marks appeared on the screen during the test interval. In both recognition conditions, the three correct recall items and three distractors were presented on the screen during the test interval in an array of 2 Columns × 3 Rows. The *S*s were required to select the correct items and write them in the correct order. In the forced-choice recognition procedure with old distractors (OD), *S* had to select and order the correct items when the distractors were the three retention items from the immediately preceding trial. In the other forced-choice recognition condition, the distractors were new (ND); i.e., they were words which had not appeared in previous trials. The set of words that were not being remembered was used as a set of distractors for the ND condition.

TABLE 1

ORDERED TRIAD SETS WHICH SERVED AS RECALL ITEMS

Set 1	Set 2
back quit disease	impossible lost danger
debt trouble out	throw missing cut
terrible foreign cost	enemy bad refuse
fear burn take	court thing hate
shoot difficult break	problem fell worry

Experimental design.—The factorial combination of the three memory test conditions on the first four trials with the test conditions on the fifth trial resulted in nine conditions assigned to nine different groups of *S*s. On the first four trials, one-third of the *S*s used the recall procedure, one-third used ND, and one-third used OD. On Trial 5, these groups were further subdivided so that one-third of each group used one of the three memory test procedures. In this way, Trial 5 performance was assessed under each of the three memory test procedures following four trials under each procedure. Each *S* experienced five different word triads on the five trials. Five orders of assignment of word triads to trials were used so that each triad occurred once on each trial over five *S*s. A complete replication consisted of five *S*s in each of nine groups. Three replications, which differed in the sets used as recall items and the positions of targets and distractors in the recognition arrays, were used.

Apparatus.—Instructions were played on a cassette tape recorder. Important aspects of the task were illustrated on cards which *E* showed to *S* at the appropriate time during the instruction period. Materials were visually presented by a Kodak Ektagraphic slide projector driven by a Sony stereo tape recorder working through a Kodak Sound Synchronizer. Pulses for the recording were provided by electronic timing equipment. A tone was provided by a Mallory Sonalert, also operated by the timing equipment, and played back on the second stereo channel through a speaker.

The experiment was run in one room of a trailer parked on a relatively quiet section of the main concourse of the Central Canada Exhibition. Another experiment was run concurrently in the second room of the trailer. Noise from the other room and from crowds outside the trailer was audible, but generally was not understandable. The equipment and an exhaust fan provided the dominant noise and had the beneficial effect of partially masking more potentially distracting noises. There was no indication that *S*s were distracted from their task and it is unlikely that the detectable noise differentially influenced the results obtained.

Procedure.—After instructions, *S* had eight trials of practice on the interpolated digit classification task alone. This was followed by four complete trials which illustrated the recall procedure (Practice

Trials 1 and 3) and ND recognition procedure (Trials 2 and 4) using consonant triads as recall material.

All experimental trials began with a brief tone sounding and three asterisks appearing on the screen for 2 sec. The recall items were then shown for 1.2 sec. while *S* read them aloud. The digit slide appeared for 16 sec. while *S* performed as many number classifications as possible. As soon as the digit slide was removed the tone sounded and one of the three types of memory test slides appeared for 15 sec. During the test interval, *S* recorded his answers on one page of a mimeographed booklet. A tone sounded and the question marks or recognition array were removed from the screen at the end of the test interval. A 4-sec. interval, in which *S* turned the page with his answers under the booklet, followed the test interval.

RESULTS

An item recalled was scored as correct if it was the correct item or an obvious misspelling of the correct item. Although *S*s were requested to name three words on every trial, guessing if necessary, there were 13% omissions for the recall condition, 3% for the ND condition, and 1% for the OD condition.

Preshift trials.—All dependent measures were subjected to analyses of variance with memory test conditions and replications as between-*S* variables and Trials 1–4 as a within-*S* variable.

The dependent measures analyzed were (*a*) item scores, the number of words retained on each trial regardless of their serial position; (*b*) order scores, the number of words recalled in the correct serial position; (*c*) a measure which adds item and order scores (Wright, 1967); and (*d*) an index identical to item scores except that 1 point is added to the item score if the words are all in the correct serial position. In addition, because there was heterogeneity of variance for item scores, a transformation was performed using the procedure described by Mendenhall (1968, p. 207). All item scores were subtracted from 3 to obtain forgetting scores, and the means were plotted against variances. Fitting a power function to these values indicated that each forgetting score should be raised to the .6 power. The same pattern of means, significant *F* ratios, and individual comparisons were found for item scores, transformed item scores, the Wright (1967) index and the short-term-memory (STM) index, while results for order scores were different. Results are presented for item and order scores.

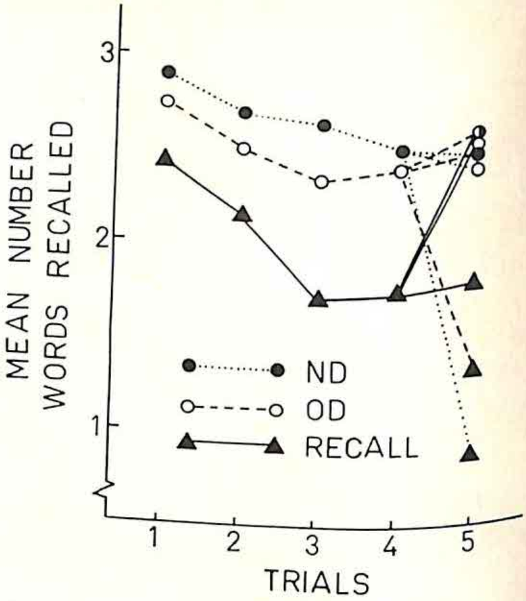


FIG. 1. Mean number of words retained regardless of order for recall, recognition with old distractors (OD), and recognition with new distractors (ND). (Each point for Trials 1–4 is based on 45 *S*s and each point for Trial 5 is based on 15 *S*s.)

Mean item scores for the three memory test conditions over the first four trials are shown in Fig. 1.

Planned orthogonal comparisons revealed that the difference between Cond. ND and OD was not significant, $F\ (1, 126) = 2.70$, $p > .10$, while the difference between recall and the average of ND and OD was, $F\ (1, 126) = 49.53$, $p < .001$. The trials effect, when tested with conservative degrees of freedom (Kirk, 1968), was significant, $F\ (1, 126) = 13.70$, $p < .001$.

The significant main effects of trials and conditions and the failure of the Conditions × Trials interaction to reach significance with substantial statistical power, $F\ (6, 378) = 1.40$, $p > .20$, taken together, indicate that forgetting increased over trials in the recognition-memory conditions and that the rates of increase were similar for Cond. ND, OD, and recall. Differences between recall and recognition occurred on the first trial, but the three procedures showed similar relative decrements over trials. When recognition scores were corrected by the traditional method (Kintsch, 1970, p. 342) for guessing, the three curves became even closer to paralleling each other and the superiority of recognition to recall was attenuated considerably.

There were reliable differences among the three replications, $F\ (2, 126) = 3.18$, $p < .05$, with a mean of 2.40 and 2.52 words retained from Set 1 and 2.20 words retained from Set 2.

Since the distinction between Cond. OD and ND on Trial 1 is meaningless because both conditions used exactly the same test items and recognition arrays, additional analyses of variance were performed on raw and transformed item scores using conditions, replications, and Trials 2–4 as variables. Results for both types of scores were essentially the same as the previous analysis. For untransformed item scores, the decrement in item availability over trials was not significant by conservative test, $F\ (1, 126) = 3.70$, $.05 < p < .10$, but was significant with conventional degrees of freedom. When transformed scores were examined, this ambiguity was resolved since the decrement was significant by both tests, $F\ (1, 126) = 4.16$, $p < .05$. The Trials × Conditions interaction was not significant, $F\ (4, 252) = 1.14$, $p > .30$, for untransformed scores. Planned orthogonal comparisons revealed that the two recognition conditions did not differ from each other, $F\ (1, 126) = 2.54$, $p > .10$, but that the two recognition conditions when combined differed from the recall condition, $F\ (1, 126) = 42.75$, $p < .001$.

Analysis of variance of order scores for Trials 1–4 revealed only a significant trials effect; by conservative test, $F\ (1, 126) = 22.54$, $p < .001$. Memory test conditions did not differ significantly in ordered recall ($F < 1$) in spite of the fact that items were differentially available for the three conditions. Since there were less items available in the recall condition, it might be expected that there would be less items in the correct intratriad position for recall. This was not the case since, of the words that were available, more were retained in the correct position for the recall condition than for the recognition conditions. Figure 2 shows the probability that a word was retained in the correct position, given that it was available. These probabilities were obtained by dividing the total number of words in the correct position by the total number of words available over all *S*s. The conditional probabilities for the recall condition are sufficiently higher than for the recognition conditions to offset the lower availability of words. Figure 2 also shows that, even when items were available, there was a decrease in position information over the first four trials.

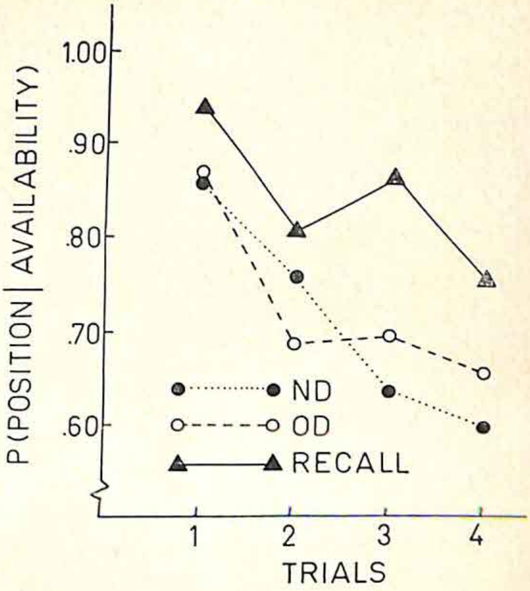


FIG. 2. Probabilities of retaining correct position information for items that were available.

Shift trials.—Analyses of variance were performed on all dependent measures with the nine Trial 5 conditions and replications as between-*S* variables and Trials 4 and 5 as a within-*S* variable. The main effect of item scores for the nine conditions shown in Fig. 1 was significant, $F\ (8, 108) = 5.39$, $p < .001$, and these conditions interacted significantly with trials, $F\ (8, 108) = 7.80$, $p < .001$. There were no other significant effects. Tests for differences using the Newman-Keuls procedure with $\alpha = .05$ were made among means for the nine conditions on Trial 5. Performance on the OD and ND conditions on Trial 5 was independent of the memory test conditions on Trials 1–4 from which they had been shifted. No significant differences were found among the three OD and the three ND conditions on Trial 5 and the three OD conditions did not differ from the ND conditions.

Performance for recall on Trial 5 did depend upon the test condition experienced on Trials 1–4. Trial 4 and 5 performance did not differ for the recall condition. When recall followed Cond. ND, recall performance was significantly worse (.87 items recalled) than when recall followed recall (1.80 items recalled), but this difference must be interpreted with caution since it did not attain significance when transformed scores were analyzed, nor with a more conservative Scheffé test. Although recall following Cond. OD (1.33 items recalled) was lower than under continuous recall, this difference was not significant. Means for the two recall groups shifted to recall from recognition differed significantly from the means for all recognition groups on Trial 5, while performance of the group maintained on recall did not differ significantly from any recognition conditions on Trial 5. However, comparable differences between recall and recognition on Trial 4 were significant with 45 *S*s in each group.

As would be expected from performance on preshift trials, Trial 5 order scores did not differ among any of the nine conditions. The only significant effect was due to replications, $F\ (2, 108) = 4.79, p < .05$. Mean number of words retained for the two replications which used recall items from Set 1 were 1.46 and 1.80, while 1.28 words were remembered for the replication which used Set 2.

The conditional probabilities of retaining words in the correct position given that they were available were based on so few words available on Trial 5 that they were unstable and no systematic results were obtained.

DISCUSSION

For all three memory test conditions, there was a loss of item availability and, in addition, a loss of order information over trials. Conditional probabilities for position information comparable to those presented here for the recall condition have been found in data from other experiments in our laboratory. Bolton (1892) and McNicol (1970) have argued that order/position information is lost faster than item availability. Murdock and Vom Saal (1967) have suggested that item availability depends to some extent on position information. The nature of such a dependency is complex because the loss of position information shown in Fig. 2 is based only on items that were available.

Proactive interference developed in the two recognition conditions at a rate which did not differ reliably from the rate for the recall condition. Any commonly used correction for guessing would make the rates of PI buildup even more similar for recall and recognition. Contrary to the results of many studies (e.g., Wickens, 1970) performance on the first trial for the recall condition was less than perfect. Studies that find less than perfect recall on the first trial are now beginning to appear in the literature (e.g., Baddeley & Scott, 1971; Goggin & Wickens, 1971). The low level of recall on the initial trial is in opposition to the view that PI is the only or even the major cause of forgetting in the Brown-Peterson paradigm.

The short presentation time could possibly account for poor recall on the first trial, but recognition performance indicates that first trial items were readily perceived and were encoded. In addition, the decrement in performance for all conditions over trials is probably not unique to a short presentation time, although, obviously, very long presentations would result in little forgetting. If *S*s were surprised by the short presentation and had trouble encoding words on Trial 1, it seems likely that they would be more prepared for the short presentation time on subsequent trials and would make a greater effort to encode them. If this happened, the encoding would be better over trials, not worse. In short, it seems likely that poor performance on Trial 1 is, to some extent, dependent upon a short presentation time, but that PI is not an artifact of the short presentation time.

Hawkins et al. (1972), using materials and procedures which differed considerably from those reported here, found a difference between recognition with old distractors and with new distractors. In their experiment, recognition on Trial 2 following recall on Trial 1 was better with new distractors than with old distractors. In view of these results and in view of the theoretical explanation for PI offered by Keppel and Underwood (1962), it was surprising that recognition performance did not differ reliably for the two types of distractors. According to the Keppel and Underwood explanation of PI, the recognition array for Cond. OD provided the most potent source of interference, previous items and present items mixed together during the memory test period. The difference in performance on Cond. OD and ND on Trial 1 is an artifact, since the two conditions were exactly the same there, and this difference was maintained for the next two trials. On Trials 4 and 5, performance on Cond. ND and OD did not differ. Furthermore, performance for Cond. ND, where the presumably competing responses were explicit, was substantially better than for recall, where presumably the same competing responses were implicit. The sizable difference in performance between recall and Cond. OD on Trial 4 together with the failure to find a significant difference between Cond. OD and ND question the importance of competition from previous trials at the time of retention test as a source of forgetting in the Brown-Peterson situation.

There is support for this conclusion if the commonly held view that a recognition procedure bypasses retrieval processes (Kintsch, 1970; McCormack, 1972) is true. Proactive inhibition, according to Keppel and Underwood (1962), is presumably a retrieval phenomenon. The demonstration of PI in recognition memory suggests that PI takes place at encoding and/or storage rather than at retrieval.

In general, performance on the shift trial did not depend on preshift test conditions. Performance on four trials under the recall condition did not lower recognition on Trial 5. Similarly, the high level of performance under the two recognition conditions for the first four trials did not raise the level of recall performance on Trial 5. Recall following recognition was at least as poor as, and perhaps poorer than, recall following recall. The demonstration of PI under all test conditions and the relative independence of PI from previous test history is consistent with other results. Cermak (1970), in a design similar to this one, varied intertrial interval and found that recall following any interval depended on the length of that interval but was not influenced by the length of intertrial interval on previous trials. It seems that PI in the Brown-Peterson situation depends upon the presence of previous trials for forgetting, but the nature of those trials, other than the type of recall material, is unimportant.

The results for recall on the shift trial are difficult to interpret in view of the dramatic results with small samples. One possibility—implicit in the previous conclusion that Trial 5 conditions were independent of preshift test conditions—is that the differences are not reliable. On the other hand, if the differences are real, they could be quite informative. The poor recall following recognition could indicate that items are not encoded and stored as well under recognition conditions (Loftus, 1971). Two aspects of the data are not consistent with this view. If items are encoded less well under recognition test conditions, various types of information, e.g., retention of the position of words, should get worse over trials relative to the recall condition. On Trial 1, when *S*s did not know which test condition would be used, there were differences in position information, but the decrement over trials was comparable for recall and recognition conditions (see Fig. 2). Furthermore, if *S*s were encoding better under recall conditions, the shift to recognition on Trial 5 should result in considerably better performance than continuous recognition. The "overshoots" when shifting from recall to recognition are considerably smaller than the "undershoots" when shifting from recognition to recall (see Fig. 1).

Another interpretation is that the different overshoots for recall after recognition with old and new distractors reflect differences in PI which cannot be detected with a recognition task, but which have graded detrimental effects on recall. The *S*s were exposed to twice as many items in Cond. ND as in OD. The differences on Trial 5 recall may reflect differences associated with interference from distractor items. This suggests that the occurrence of different items from the same encoding class is potentially a more important source of interference than repetitions of previous targets.

REFERENCES

Baddeley, A. D., & Scott, D. D. Short-term forgetting in the absence of proactive interference. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 275–283.

Bolton, T. H. The growth of memory in school children. *American Journal of Psychology*, 1892, **4**, 362–380.

Cermak, L. S. Decay of interference as a function of the intertrial interval in short-term memory. *Journal of Experimental Psychology*, 1970, **84**, 499–501.

Goggin, J., & Wickens, D. D. Proactive interference and language change in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 453–458.

Hawkins, H. L., Pardo, V. J., & Cox, R. D. Proactive interference in short-term recognition: Trace interaction or competition? *Journal of Experimental Psychology*, 1972, **92**, 43–48.

Heise, D. R. Semantic differential profiles for 1,000 most frequent English words. *Psychological Monographs*, 1965, **79** (8, Whole No. 601).

Keppel, G., & Underwood, B. J. Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 153–161.

Kintsch, W. Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

Kirk, R. E. *Experimental design: Procedures for the behavioral sciences*. Belmont, Cal.: Wadsworth, 1968.

Loftus, G. R. Comparison of recognition and recall in a continuous memory task. *Journal of Experimental Psychology*, 1971, **91**, 220–226.

McCormack, P. D. Recognition memory: How complex a retrieval system? *Canadian Journal of Psychology*, 1972, **26**, 19–41.

McNicol, D. Progressive decay of the memory trace. *Psychonomic Science*, 1970, **20**, 196.

Mendenhall, W. *Introduction to linear models and the design and analysis of experiments*. Belmont, Cal.: Wadsworth, 1968.

Murdock, B. B., Jr. The retention of individual items. *Journal of Experimental Psychology*, 1961, **62**, 618–625.

Murdock, B. B., Jr., & Vom Saal, W. Transpositions in short-term memory. *Journal of Experimental Psychology*, 1967, **74**, 137–143.

Wickens, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

Wright, J. H. Effects of formal interitem similarity and length of retention interval on proactive inhibition of short-term memory. *Journal of Experimental Psychology*, 1967, **75**, 386–395.

(Received March 6, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 419–428

FREE RECALL OF GROUPED WORDS [1]

ROSAMOND GIANUTSOS [2]

New York University

Grouped and ungrouped lists were presented for free recall. Temporal grouping increased recall of only items presented at the end of the list, i.e., short-term storage. Contrary to expectations, grouping did not improve long-term storage. The fact that grouping produced a stereotypic order of recall focused on recall order as a crucial factor in determining short-term recall.

Grouping is interesting not only because it is used frequently as a mnemonic technique but also because it is intimately involved in speech in the form of intonation. Intonational groupings (created by pausing or accenting) play an important role in processing and remembering verbal stimuli (O'Connell, Turner, & Onuska, 1968). The purpose of the present series of studies was to elucidate the role of grouping in remembering words by using single-trial free recall. The U-shaped serial position curve obtained with free recall is viewed as a composite with recall for items from the end of the list being determined primarily by the state of the short-term store (STS) and recall for other items being determined by the long-term store (LTS) (Atkinson & Shiffrin, 1968; Glanzer, 1968; Waugh & Norman, 1965).

Previous studies of rhythmic or temporal grouping have involved fixed-order recall of items such as numbers, letters, and nonsense syllables. Ryan (1967) reviewed these studies. However, fixed-order recall is in two ways a more complex task than free recall: (*a*) It requires order information; and (*b*) it requires a fixed order in recalling words. These factors make it difficult to analyze the role of STS and LTS processes, as can be done with free recall.

The previous studies of grouping generally found that temporal grouping improved recall. Ryan (1967) showed that this improvement was due primarily to an increase in order information rather than to an increase in item information. Because the recall of order information is not required in free recall, the improvement found in fixed order recall would not necessarily be expected in free recall.

EXPERIMENT I

This experiment was designed to determine the effects of temporal grouping on LTS and STS. Because temporal grouping involves the establishment of a rhythm, it was hypothesized that overall presentation rate might be important. Therefore, total list presentation time was varied from 14.5 sec. to 24 sec. for 18-word lists. Control and experimental lists were equated on the basis of total presentation time. However, the experimental design also permitted comparison of conditions at the same absolute presentation rate.

Method

Apparatus.—A memory drum specially designed to receive computer listings was used to present the stimuli. All of *S*'s responses were tape recorded. A computer was used to produce all randomizations of the stimulus materials for each *S*, in addition to printing the memory drum "tapes." The same apparatus was used in later experiments, unless otherwise indicated.

Subjects.—Fifty introductory psychology students from New York University participated as a course requirement. None had previous experience in a free-recall experiment. The *S*s from the same source were used in the later experiments, but no *S* was used in more than one of these experiments.

---

[1] The author thanks both her sponsor, Murray Glanzer, and Doris Aaronson for extensive guidance and encouragement. Reinhilde Nelson was most helpful in testing *S*s. This paper is based on a dissertation submitted in partial fulfillment of the requirements for the PhD degree at New York University. While conducting this research the author was a National Institutes of Health predoctoral fellow (5 F01 MH40050).

[2] Requests for reprints should be sent to Rosamond Gianutsos, who is now at the Department of Psychology, Adelphi University, Garden City, Long Island, New York 11530.

Materials.—Lists consisted of high-frequency (AA, Thorndike & Lorge) monosyllabic nouns.

Procedure.—Each *S* was presented with 12, 18-word lists for free recall. The *S* read the words aloud as they appeared. The end of the list was indicated by a blank space colored yellow. The recall period was 1 min. The two main variables in this experiment were total list presentation time and temporal grouping of stimulus words. Grouped lists had a blank space after every third word. It was decided to use groups of 3 words because Ryan (1967) found that grouping effects were most marked for regular groups of 3.

Because grouped lists contained blank spaces, it was necessary to match grouped and ungrouped lists on the basis of total list presentation time. The importance of total time has been shown by a number of investigators, including Murdock (1960). Three presentation times were used: 14.5, 19, and 24 sec. The time for each word (including time for drum advancement) ranged from .6 sec. in the grouped 14.5-sec. lists to 1.3 sec. in the ungrouped 24-sec. lists. The pairs of these conditions were also matched for exposure time per item: the ungrouped 14.5-sec. and the grouped 19-sec. lists; and the ungrouped 19-sec. and the grouped 24-sec. lists.

Each *S* was tested individually. He was first given two practice lists: an ungrouped 24-sec. list followed by a grouped 14.5-sec. list. In the experimental lists, all six condition combinations (2 Levels of Grouping × 3 Levels of Presentation Time) were represented twice. The presentation order of conditions was randomized independently for each *S*, as was the assignment of words to lists.

Results

Probability of recall.—The reliability of the differences illustrated in Fig. 1 was assessed by analysis of variance. The analysis of the list as a whole was supplemented by separate analyses of the individual groups, as detailed by Gianutsos (1970).

Grouping depressed recall in Group 1 at the beginning of the list, $F\ (1, 49) = 7.68$, and raised recall for the two groups at the end of the list, Fs $(1, 49) = 23.38$ and 12.52, all ps $< .01$. These effects offset each other so that in the analysis of the list as a whole grouping was not significant, $F\ (1, 49) = 1.39$, $p > .05$; however, there was a significant interaction of grouping with serial position, $F\ (17, 833) = 3.77$, $p < .01$. The increase in recall in the last group was largest for the first item of the group. This interaction (Grouping × Serial Position) was significant, $F\ (2, 833) = 7.91$, $p < .01$.

As for the presentation time variable, there was an improvement in recall for the slower lists, primarily in the first two thirds of the list. In the list as a whole, presentation time, $F\ (2, 98) = 7.79$, serial position, $F\ (17, 833) = 206.46$, and their interaction, $F\ (34, 1{,}666) = 2.02$, all ps $< .01$, were all significant. These effects were to be expected from previous research (Glanzer & Cunitz, 1966; Raymond, 1969). The interaction reflects the fact that the elevation of recall found in slower lists was most pronounced in the first part of the list. In the analyses of the individual groups,

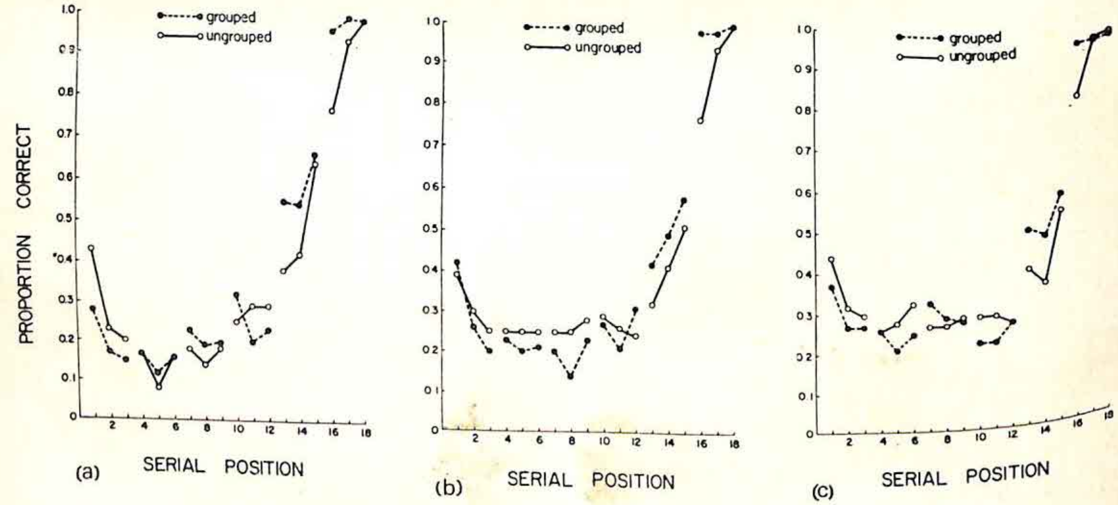


FIG. 1. Temporal Grouping 1: Serial position functions for temporally grouped and ungrouped lists. (The total presentation times and intervals between groups were (*a*) 14.5 sec. and .6 sec.; (*b*) 19 sec. and .8 sec.; and (*c*) 24 sec. and 1.0 sec.)

one isolated significant interaction with presentation time occurred. This was in the third group, $F\ (2, 98) = 3.31, p < .05$, but the effect was not a systematic one; it did not recur in later experiments, and it is not relevant to the present theoretical approach.

Order of recall.—Grouping produced a highly predictable order of recall. The last group, and often the next-to-last group, were recalled in order. Typically the last group was recalled first, often followed by the next-to-last group. Less frequently, the next-to-last group would be recalled first, followed by the last group. With ungrouped lists there was no regular pattern except that, as Deese and Kaufman (1957) observed, recall usually began with items from the last part of the list followed by items from the beginning. In grouped lists recall began consistently with the first word in the last group 83% of the time, whereas in ungrouped lists recall was most likely to begin with the last word in the list, which, however, was chosen only 35% of the time, reflecting a less consistent patterning of recall order for ungrouped lists.

Discussion

The results of the first experiment show that temporal grouping is beneficial to STS. Improvement was particularly noticeable for the first word in each of the last two groups. Recall for the last three words (one group) of grouped lists was nearly perfect; while for ungrouped lists nearly perfect recall was observed for only the very last word in the list. This fact gave the serial position curves for grouped lists a "steplike" appearance at the end of the list; that is, the grouped serial position curve resembled the profile of a staircase. No other interactions were found.

How does temporal grouping produce the discontinuous steplike increment in recall? The discussion of this question will have two parts: The first part is a demonstration that the changed order of recall found in grouped lists produces a serial position function with steplike increments. The second part is an explanation of why temporal grouping changes recall order.

The effect of recall order on STS.—Whether items are lost from STS because of temporal decay (Brown, 1958; Conrad & Hille, 1958) or because of displacement by subsequent items (Glanzer, Gianutsos, & Dubin, 1969), the likelihood of recall is dependent upon the order of recall (Bousfield, Cohen, & Silva, 1956; Tulving & Arbuckle, 1963). The displacement model will be used in further discussions, together with the following two assumptions: (*a*) Words read as stimuli and words given as responses both displace previous items in STS (Tulving & Arbuckle, 1963; Waugh & Norman, 1965). (*b*) Each new item tends to displace the oldest one in STS. Recall of any item reintroduces that item into the STS as if it were "new," displacing the oldest one out. Accordingly, the approach of recalling backwards through the list is inefficient because it displaces items before they are recalled. In order to recall the most items from STS, the optimal strategy would be to recall the word that is about to be displaced, the oldest word in the store. If STS contains words originally in Positions 15, 16, 17, 18 at the end of the list, one should start by recalling 15, i.e., so that store will then contain 16, 17, 18, 15. Then 16 should be recalled, and so on until all the items in STS at the end of list presentation are recalled. This sort of forward-ordered recall within the span of STS is what is increased by temporal grouping.

According to Gianutsos (1970), the shape of the short-term component of the serial position curve depends on the order in which items are recalled. With ungrouped lists unsystematic orders of recall predominate, although recall begins most often with the last item in the list. Because of the irregular order of recall, the short-term component of ungrouped lists is expected to increase smoothly. This smoothness is produced by averaging. On the other hand, for grouped lists a modified forward-ordered recall predominates with recall proceeding forward from the first member of the later groups. This type of recall order results in a short-term serial position curve that increases in a steplike fashion.

The effect of grouping on recall order.—The reason that temporal grouping changes the order of recall, argued for here, is that when *Ss* read a grouped list they tend to impose an intonation pattern on the items. Such prosodic features may be retained with items in STS. It would be inefficient to dissociate them in spoken recall from STS. In other words, it would be easier to recall items from STS in the intonation pattern in which they were presented than in some other intonation. Since intonation patterns must be reproduced in order, recall that reflects intonation patterns will be in order. In short, it is to preserve intonation that forward-ordered recall is used with temporally grouped lists. On the other hand, while it is advantageous to recall items within groups in order, there is no special advantage—as far as preserving intonation cues—to recalling the next-to-last group before the last group. This finding contradicts an interpretation of grouping by Wickelgren (1967), who implies that the groups, as well as the items within groups, should be recalled in order.

EXPERIMENT II

This experiment served primarily to replicate and to extend the range of the variables in Exp. I. Since recall for the early parts of the list was relatively low in Exp. I, in Exp. II the total list presentation time was increased to 36 sec. In the past (Raymond, 1969) slower rates have been found to raise the level of LTS in free recall. In Exp. II, the additional time was to make sure that any effects on LTS were not obscured by "floor" effects.

Method

Subjects.—Sixty college students participated.

Materials.—The stimulus words and the manner of randomizing their presentation were identical to those used in Exp. I.

Procedure.—This experiment was similar to the preceding one with the following exceptions: (*a*) There were three levels of temporal grouping. In addition to ungrouped lists and lists with one blank item between groups of three words, there were also lists with two blank items between groups. (*b*) Grouping conditions were presented in blocks. Each of the six different orders of the three types of blocks was used. Five *S*s were assigned to each block order. Blocking was introduced so that any transfer of experience from one level of grouping to another could be measured. In the preceding experiment it had been observed that after *S*s experienced grouped lists, they tended to impose intonational grouping on ungrouped lists. (*c*) There were two total list presentation times: 18 sec. and 36 sec. The exposure time for an individual item ranged from .6 sec. to 1.9 sec. The presentation rates were assigned randomly to lists within blocks. (*d*) The two practice lists were both ungrouped, the first being a 36-sec. list and the second being an 18-sec. one.

Results

Probability of recall.—As shown in Fig. 2, grouping increased recall more at the end of the list than it decreased it at the beginning. Averaged over all serial positions, the effect of grouping was significant, $F\ (2, 118) = 7.38$, $p < .01$. The effect of grouping was significant for the first and also the last two groups, Fs $(2, 118) = 4.54$, 9.56, and 11.46, ps $< .05$, .01, and .01. The interaction of grouping with serial position was significant, $F\ (34, 3{,}006) = 3.08$, $p < .01$, reflecting the two-way effect. Grouping also interacted with serial position in the last two groups, Fs $(4, 2{,}006) = 2.71$ and 5.27, ps $< .05$ and .01, reflecting the particular advantage of grouping for the first items in the groups.

The effects of presentation time again confirmed expectations: Increased presentation time significantly increased recall in Serial Positions 1–12, Fs $(1, 59) = 44.89$, 16.24, 34.11, and 27.11, ps $< .01$; however, recall was reduced significantly in the next-to-last group, $F\ (1, 59) = 5.98$, $p < .05$. There was a nonsignificant reduction in the last group as well.[3] Presentation time also interacted significantly with serial position in Groups 4 and 5, Fs $(2, 1{,}003) = 4.01$ and 4.37, ps $< .05$. These effects are not large. From the serial position curve it appears that taking more time to present a list dilutes the effects of grouping. These trends were confirmed only for Group 5 in which the grouping effect depended on presentation time, $F\ (2, 118) = 3.27$, $p < .05$. Block order was not significant either as a main effect or in its interactions with grouping, indicating that there was no significant transfer of effects between grouped and ungrouped conditions.

Order of recall.—Most of the comments on recall order made about Exp. I hold for Exp. II. Again, recall began consistently with the first item in the last group in 73% of the grouped lists, whereas for ungrouped lists the first recalled item was most likely to be the last item in the list, but this

---

[3] This two-way effect may be the cause of the conflicting literature on the effect of rate on recall: In some studies increased rate has been found to be beneficial and in others deterimental to recall. See Aaronson (1967) for a review of this conflict. The present findings suggest that fast rates may benefit STS by inducing grouping, while being detrimental to LTS.

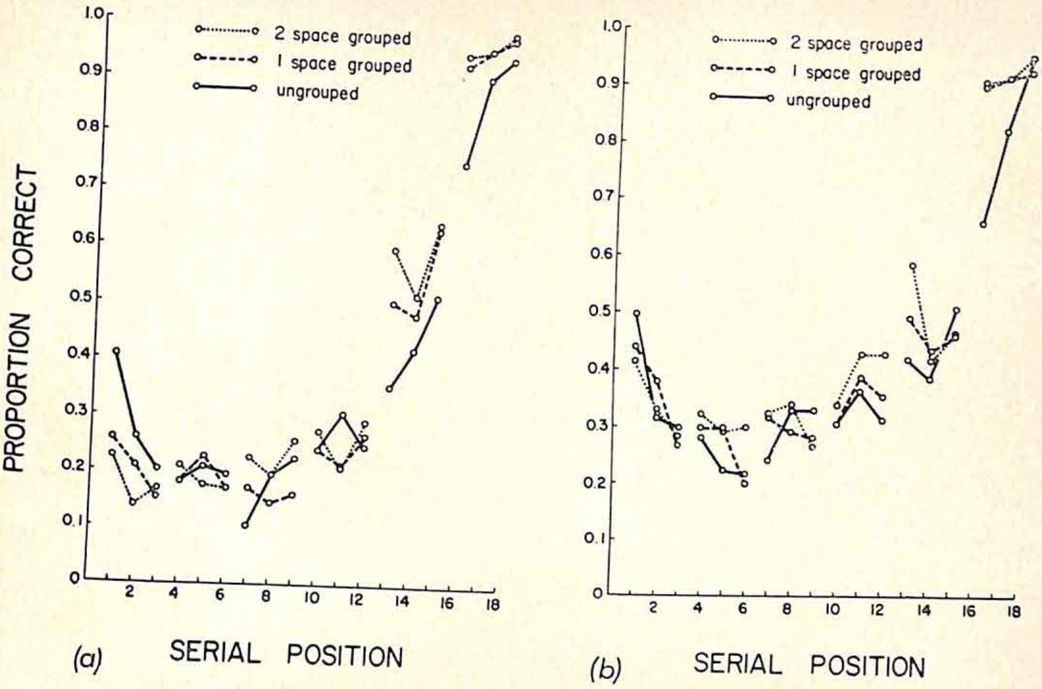


FIG. 2. Temporal Grouping 2: Serial position functions for three conditions of temporal grouping are 2-space grouped, 1-space grouped, and ungrouped. (The total presentation times and intervals between groups in the two grouping conditions were (*a*) 18 sec., 1.2 sec., and .8 sec. and (*b*) 36 sec., 2.5 sec., and 1.5 sec.)

tendency was less general, happening in 42% of the lists. Considering sequences involving only the last six words in the list, the proportion of recalled pairs that were forward ordered was .54 for ungrouped lists and .72 for grouped lists.[4] Both types of 18-sec. lists were recalled with a more consistent selection of the typical recall order (derived from maximum transitional probabilities as detailed by Gianutsos, 1970) than the 36-sec. lists. Moreover, the two-space grouped condition was recalled with a more consistent selection of the typical recall order than the one-space grouped condition.

Discussion

Degree of grouping.—Where the effects of grouping appeared, an increase in the intergroup interval led to an increased grouping effect. For STS, it is consistent with the model proposed in the discussion of Exp. I that a longer intergroup interval should increase the grouping effects. The longer interval leads to a more accentuated grouping intonation pattern. This could be expected to produce an even more consistent use of the typical grouping order of recall, and, in fact, it did. Consequently, the discontinuity of the increment in STS recall would be enhanced.

Presentation time.—The effects of grouping on STS were reduced by the increased presentation time. If the rate of presentation is slowed too much, it is possible that the intonation pattern or breath group would be changed. A single intonation or breath group would be associated with a group of stimulus words only if the presentation time for the group does not exceed the duration of breath groups in ordinary speech. Lieberman (1967) notes that although a single expiration may endure for up to 40 sec., the normal duration of breath groups is much shorter. For example, test sentences such as "Joe ate his soup" and

---

[4] For ungrouped lists, the typical recall order based on maximum transition probabilities was simply 18 and 17, then recycling back to 18. For grouped lists, on the other hand, it was the following: 16, 17, 18, 13, 14, 15, recycled at this point into Item 16—although in the 36-sec., two-space grouped condition it was equally likely that recall would end with Item 15. This pattern of recall amounts to recalling the last group in order first, followed by the next-to-last group in order. The fact that the "typical recall order" is much longer for grouped lists is indicative of the more extensive regularities in recall order to be found among lists.

"You're going to drive down that rutted road" are spoken in 1.5 to 2 sec. In the 18-sec. lists of Exp. II each group of three words was presented for a total of 2.4 and 1.8 sec. for the one-space and two-space grouped conditions, respectively, not counting the intergroup interval. In contrast, the comparable times for the 36-sec. lists are 4.5 and 3.6 sec. These data suggest that the longer presentation time extended the duration of the stimulus group to exceed that of the normal-breath group. It is possible that grouping might have been more effective if *S*s were encouraged to read all the words in each group in a single breath.

In conclusion, the effects of temporal grouping are enhanced by increasing the interval between groups and are greater for shorter presentation times.

EXPERIMENT III

The purpose of Exp. III was to simulate the important aspects of grouping using a method other than temporal grouping. It was hypothesized that the important aspects of temporal grouping are (*a*) that it changes recall order and (*b*) that it divides the list into smaller segments or "chunks." In this experiment, therefore, some *S*s were instructed to use the order of recall typically used for temporal grouping. In addition, nontemporal grouping was used, instead of temporal grouping, to divide the list into smaller units. If these variables produce the same results as temporal grouping, then the effect of the pause between temporally defined groups can be viewed as simply a marker, setting off the groups.

Method

Subjects.—Forty college students participated.

Materials.—The stimulus words and the method of randomizing their presentation order were identical to those used in the two preceding experiments.

Procedure.—This experiment was generally similar to the preceding experiments. Grouping was again a variable, but here the groups were indicated by coloring the background of the words appropriately. They were not temporally demarcated. The words were presented at evenly spaced intervals, with every other group of three words colored yellow to indicate grouping.

A second variable was instructions as to recall order. The specially instructed *S*s were told to view the list as consecutive groups of three words (as indicated by color, when colored groups were used). They were told to recall the words within a group in order, starting with the last group, then the next-to-last group, etc. The *S*s were told to abandon this order of recall for a list when it became difficult to maintain. They did, in fact, abandon it after giving the last or the last two groups.

Four groups of 10 *S*s each were given one of the four possible combinations of the two levels of the nontemporal grouping and instructions variables. The following detail should be noted: *S*s who had special instructions as to recall order, but who did not have groups marked by color, tended to impose grouping cues by intonation as they read the list. When *E* observed this, she told *S* not to use intonation for grouping.

Results

Probability of recall.—The serial position curves in Fig. 3 show that color grouping tended to reduce recall at the beginning of the list and increase it slightly at the end. Since these trends cancel each other, the effect of grouping in the list as a whole was not significant but the interaction with serial position was, F (17, 153) = 2.23, $p < .01$. Grouping interacted with serial position in the fifth group, F (2, 153) = 3.24, $p < .05$, reflecting a tendency toward discontinuous steplike increases.

The effect of special instructions alone were similar to grouping, though more accentuated. Instructions reliably reduced recall in Group 1, F (1, 36) = 8.17, $p < .01$, and increased recall in the next-to-last group, F (1, 36) = 8.70, $p < .01$. The trend toward increased recall in the last group did not reach significance. The two-way effect was also shown by a significant interaction of serial position with instructions, F (17, 153) = 5.34, $p < .01$.

In combination with grouping, special instructions produced a marked decrement at the beginning of the list and distinct discontinuous increments corresponding to groups at the end. The reliability of these trends is supported by significant triple interactions (Grouping × Instructions × Serial Position) in Groups 1 and 5: Fs (2, 153) = 6.90 and 3.47, ps < .01 and .05. While the discontinuous increments at the end of the list, found previously with temporal grouping, do not appear with color grouping or instructions alone, marked discontinuous

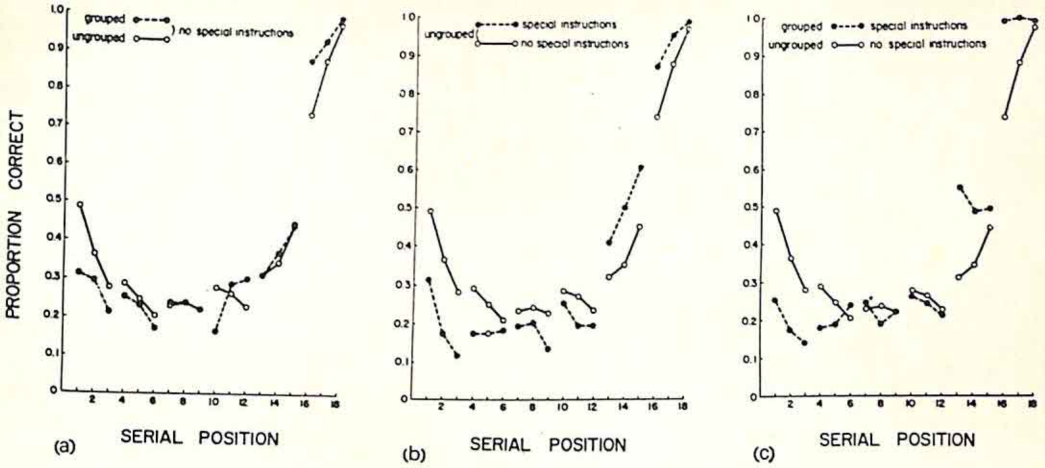


FIG. 3. Nontemporal grouping and special instructions: Serial position functions for lists with a total presentation time of 18 sec. (In all panels an experimental condition is compared with an ungrouped no-special-instruction control condition. The experimental conditions represented are (*a*) nontemporal grouping without special instructions, (*b*) special instructions without nontemporal grouping, and (*c*) nontemporal grouping with special instructions.)

increments appear when they are used together.

Order of recall.—Forward ordering of recall (which is indexed indirectly by the degree to which recall begins with items other than the last) is increased by special instructions. Without special instructions, *S*s chose to start with the last word almost 50% of the time. Whereas, with special instructions and grouping together, recall began almost exclusively ($p = .95$) with the first word in the last group (Serial Position 16). Without grouping, this preference was maintained to a lesser degree ($p = .68$) and in 22% of the lists recall began with the fifteenth or seventeenth items.

Discussion

The effects of nontemporal grouping were less than temporal grouping, though in the same direction. The finding is not surprising in light of Ryan's (1969a, 1969b) difficulty in obtaining significant effects with this variable in fixed-order recall. Instructions alone had a greater effect, emphasizing the importance of what Atkinson and Shiffrin (1968) called "control processes." Finally, grouping and special instructions combined to produce the temporal grouping pattern. Neither effect alone produced the complete temporal grouping pattern of recall.

Effects on STS.—The recall orders typical of grouped lists are induced by the intonation pattern used when reading items in temporal groups. In Exp. III, forward ordering alone (produced by alternate means) produced a continuous increase in short-term recall. When marking was added by nontemporal grouping in Exp. III, the shape of the increment in recall became steplike, like that found with temporal grouping. To summarize, temporal grouping induces forward-ordered recall within the span of the short-term store and provides markers to "anchor" recall.

Effects on LTS.—It is clear from Fig. 3a, 3b, and 3c that both nontemporal grouping and special instructions decreased recall of the beginning of the list. Moreover, special instructions decreased the amount of recall more than nontemporal grouping. These findings show that the effects of temporal grouping have no necessary relation to the allowance of extra time between groups. Effects similar to those of temporal grouping are found without any variation in the temporal properties of list presentation.

The findings also rule out the assertion that the detrimental effects of temporal grouping on LTS are merely due to the more rapid rate of presentation necessary to keep total presentation time constant. In Exp. III, there is a distinct lowering of LTS recall produced by a nontemporal grouping procedure, a procedure in which *S*s are simply told that some words will be yellow and some white. The detrimental effect is even stronger when another nontemporal procedure is used, namely, the special instructions to recall groups in order.

The finding that a grouping effect can be obtained in several different ways (by various forms of nontemporal grouping as well as by different degrees of temporal grouping) suggests that the significant aspects of these procedures are not immediate functions of their surface properties. That is, some deeper feature is shared by these procedures and is reflected, perhaps differently, in their surface properties. This deeper feature may be the distinctiveness of the smaller collections of items into which the list is divided.

EXPERIMENT IV

The purpose of this experiment was to distinguish unambiguously the effects of grouping on LTS from effects on STS by interposing a delay task between list presentation and the recall period to eliminate the STS component of the serial position curve (Glanzer, Gianutsos, & Dubin, 1969).

Method

Apparatus.—The same apparatus as that used in the preceding experiments was employed. An additional timing circuit was installed for pacing the delay task. The circuit was set at 1 pulse/sec and was activated by *E* after the main list.

Subjects.—Seventy-five college students participated.

Materials.—The same set of high-frequency (AA) monosyllabic nouns used in Exp. I–III was used in the present experiment. Lower frequency (A) monosyllabic nouns were used for the delay task.

Procedure.—The method of randomization and presentation, the list length, etc., were similar to those used in Exp. I and II. Total presentation time was fixed at 18 sec., with an 18-item list. Half of the 12 lists were presented at constant rate and half were grouped with two blank spaces between groups of 3 items. Half of the lists of each type were followed by an additional six words presented two at a time on the drum with 1 sec. for each pair. The *S* read these words out loud, but was not required to recall them. The postlist words were on a yellow background to distinguish them from the main list. Grouped and ungrouped lists were presented in blocks. Block order was counterbalanced.

Results

As can be seen in Fig. 4 the postlist task eliminated any increase at the end of the list. It therefore can be assumed that recall which followed the postlist task involved LTS and not STS.

Grouping interacted with serial position in the list as a whole, F (17, 578) = 5.37, $p < .01$. Recall was reliably reduced in Group 1, F (1, 34) = 11.48, and increased in Groups 5 and 6, Fs (1, 34) = 8.29 and 15.18, all ps $< .01$. Discontinuous increases at the end of the list in the no-postlist task condition translated into significant interactions of grouping and serial position in Groups 5 and 6, Fs (2, 578) = 4.13 and 10.93, ps $< .05$ and .01. While the Grouping × Serial Position interaction in Group 4 was significant, F (2, 578) = 3.74, $p < .05$, it was not meaningful. However, when the STS component was eliminated by the postlist task, all effects of grouping were eliminated, as shown by significant interactions of grouping with delay in the last two groups, Fs (1, 34) = 27.75 and 5.10, ps $< .01$ and .05. The discontinuity of the grouping effect in the no-postlist task condition is underscored by the significant Grouping × Delay Task × Serial Position interaction, F (17,578) = 1.78, $p < .05$.

As a main effect, block order was not significant; however, it did interact significantly with grouping, F (1, 34) = 6.53, $p < .05$, and with grouping and postlist task, F (1, 34) = 8.96, $p < .01$. The interaction of block order with grouping was found mostly at the beginning of the lists where grouping was more damaging to recall for *S*s in the second half of the experimental session. While there was some transfer from grouped to ungrouped lists, there was none in the opposite direction. The overall effect of transfer therefore was to reduce the differences between grouped and ungrouped lists.

Discussion

Since it turned out that with the postlist task grouping decreased recall of the first few items only, the mechanism by which grouping affects LTS is not so general as those suggested heretofore. Also, the fact that with the postlist task grouping had no effect at the end of the list shows that the end peak is indeed a pure STS phenomenon and not due to LTS.

For grouped lists, flattening of the initial LTS peak is explained as due to incomplete

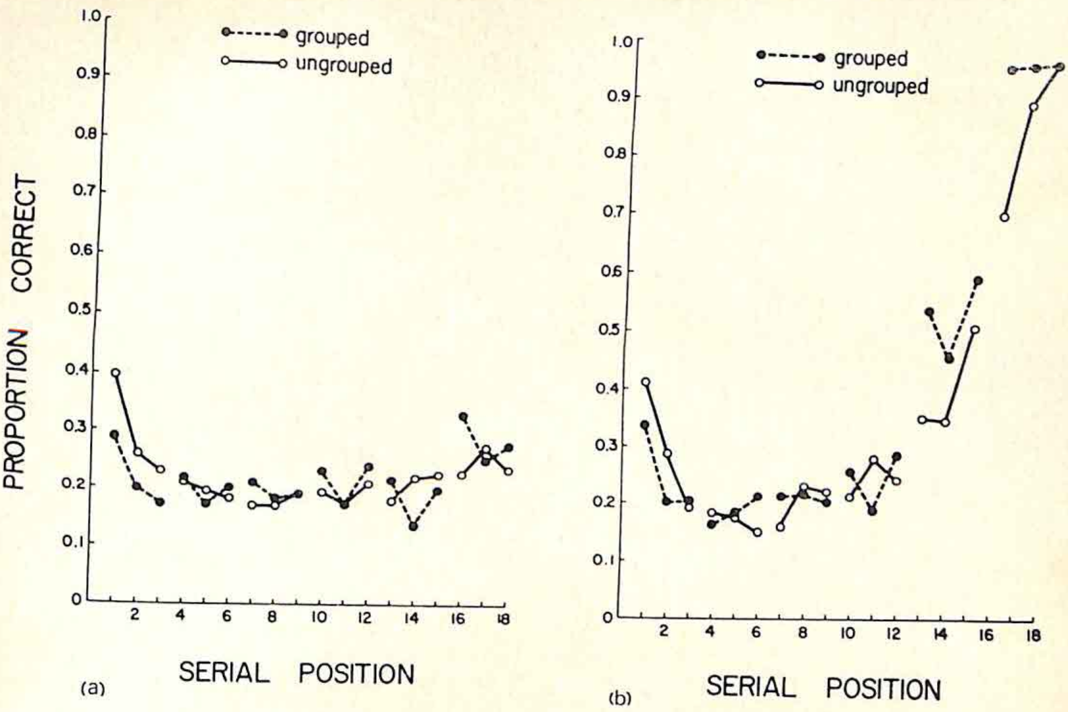


FIG. 4. Postlist task and temporal grouping: Serial position functions for temporally grouped and ungrouped lists with a total presentation time of 18 sec. and 1.2 sec. between groups. (In Panel a, a postlist task was interposed between list presentation and recall. In Panel b, no postlist task was interposed.)

processing of items at the beginning of grouped lists. Rather, much of processing is delayed until an entire group is in, either because there is less time within the group or because processing time is reallocated in accordance with some control process. Therefore, early items do not have the advantage of being in a nearly empty STS buffer at the time of processing. Ordinarily in ungrouped lists, this advantage produces the initial peak—primacy portion—of the LTS component (Bruce & Papay, 1970; Gianutsos, 1970; Rundus & Atkinson, 1970). In grouped lists less processing is done between items in the first group and more is done in the interval between Groups 1 and 2. This redistribution of processing from within to between groups is maintained throughout the list, but does not change the overall information coded for later items, since the STS is filled.

In single-trial free recall temporal grouping is beneficial only to STS. While grouping actually tends to reduce the initial peak, it does not change the asymptote of LTS. The effect of grouping on order of recall focused attention on recall order as a crucial factor in determining probability of recall from STS. The discovery that grouping affects only the initial peak or primacy portion of the serial position curve was explained in terms of the reallocation of coding time from within to between groups. In conclusion, these findings cast doubt on the value of temporal grouping as a general mnemonic device. Moreover, intonation, which includes temporal grouping, should be expected to play a role in the earlier stages of speech comprehension where STS, and not LTS, is mainly involved.

REFERENCES

AARONSON, D. Temporal factors in perception and short-term memory. *Psychological Bulletin*, 1967, **67**, 130–144.

ATKINSON, R. C., & SHIFFRIN, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. Vol. 2. New York: Academic Press, 1968.

BOUSFIELD, W. A., COHEN, B. H., & SILVA, J. G. The extension of Marbe's law to the recall of stimulus words. *American Journal of Psychology*, 1956, **69**, 429–433.

BROWN, J. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 1958, **10**, 12–21.

Bruce, D., & Papay, J. P. The primacy effect of single-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 473–486.

Conrad, R., & Hille, B. The decay theory of immediate memory and paced recall. *Canadian Journal of Psychology*, 1958, **12**, 1–6.

Deese, J., & Kaufman, R. A. Sequential effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology*, 1957, **54**, 180–187.

Gianutsos, R. Free recall of grouped words. Unpublished doctoral dissertation, New York University, 1970.

Glanzer, M. Storage mechanisms in free recall. *Transactions of the New York Academy of Sciences*, 1968, **30**, 1120–1129.

Glanzer, M., & Cunitz, A. R. Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 351–360.

Glanzer, M., Gianutsos, R., & Dubin, S. The removal of items from short-term storage. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 435–447.

Lieberman, P. *Intonation, preception, and language.* Cambridge, Mass.: The M. I. T. Press, 1967.

Murdock, B. B. The immediate retention of unrelated words. *Journal of Experimental Psychology*, 1960, **60**, 222–234.

O'Connell, D. C., Turner, E. A., & Onuska, L. A. Intonation, grammatical structure, and contextual association in immediate recall. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 110–116.

Raymond, B. Short-term storage and long-term storage in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 567–574.

Rundus, D., & Atkinson, R. C. Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 99–105.

Ryan, J. Grouping and short-term memory. Unpublished doctoral dissertation, University of Cambridge, 1967.

Ryan, J. Grouping and short-term memory: Different means and patterns of grouping. *Quarterly Journal of Experimental Psychology*, 1969, **21**, 137–147. (a)

Ryan, J. Temporal groupings, rehearsal and short-term memory. *Quarterly Journal of Experimental Psychology*, 1969, **21**, 148–155. (b)

Tulving, E., & Arbuckle, T. Y. Sources of intratrial interference in immediate recall of paired associates. *Journal of Verbal Learning and Verbal Behavior*, 1963, **1**, 321–334.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89–104.

Wickelgren, W. A. Rehearsal grouping and hierarchical organization of serial position cues in short-term memory. *Quarterly Journal of Experimental Psychology*, 1967, **19**, 97–102.

(Received March 15, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 429–436

AUDITORY CODABILITY AND THE SHORT-TERM RETENTION OF VISUAL INFORMATION[1]

R. S. NICKERSON[2]

Bolt Beranek and Newman, Inc., Cambridge, Massachusetts

The *S*'s task was to decide as quickly as possible whether the second of two sequentially presented visual stimuli was the same as the first. Three types of stimulus pairs were used: letters, random dot patterns that were easy to differentiate, and random dot patterns that were difficult to differentiate. Reaction times (RTs) were faster, and error rates lower, with the letters and easy random patterns than with the difficult random patterns. The relationship between RT and the duration of the interval between the stimuli being compared was curvilinear over the range 0 to 8 sec. The results did not support the idea that the letters, for which auditory codes presumably are readily available, were stored in short-term memory in a fundamentally different way than were the random dot patterns.

The distinction between visual and auditory (or articulatory) short-term memory (STM) is a familiar and widely accepted one. It is generally held that the contents of visual STM, or "iconic" memory as Neisser (1967) has termed it, persist only for a very brief time (a second or so) and, if they are to be retained for longer periods, must be "recoded" and represented within another storage medium. It is further held that if the visual patterns that must be remembered have highly familiar auditory associates (names), the recoding involves the substitution of these associates for the visual patterns and the retention medium becomes an auditory,[3] rather than a visual, one. Evidence for such a visual to auditory recoding comes from a variety of sources, perhaps the most compelling being experiments showing the effects of acoustic confusability on the recall of visually presented letters or digits (Conrad, 1964; Sperling & Speelman, 1970; Wickelgren, 1966). It is assumed that items in auditory STM are more easily rehearsed than those in visual STM and can, therefore, be retained for much longer periods of time; thus the recoding makes sense, given the goal of retention, and providing auditory associates of the visual patterns are easily accessed.

How visual patterns that do not have readily available verbal labels are retained in STM over intervals that exceed the persistence of iconic memory is not clear. One possibility is that they are recoded as verbal descriptions or auditorily represented feature lists (Noton & Stark, 1971). While it is apparent that some recoding of this sort does occur, it seems unlikely that one could account for our ability to remember visual stimuli exclusively on this basis. People are very good at remembering complex visual patterns of the type that would be difficult to reduce to unique or differentiating descriptions (Nickerson, 1965; Shepard, 1967).

In any case, it seems consistent with the prevailing notions about visual and auditory STM to assume that visual patterns for which there are familiar names (e.g., letters and digits) would be encoded in memory differently than would those for which familiar names do not exist. We would expect memory for items in the former category to show the characteristics of auditory STM, whereas memory for those in the latter should have more of the characteristics of a visual representation.

The results of two recent studies are, in some particulars, difficult to reconcile with this view. Nickerson and Pew (1972) had

---

[1] This experiment was supported by the United States Air Force Office of Scientific Research under Contract F44620-69-C-0115. Barbara Noel assisted in collecting and analyzing the data.

[2] Requests for reprints should be sent to Raymond S. Nickerson, Bolt Beranek and Newman, Inc., 50 Moulton Street, Cambridge, Mass. 02138.

[3] The term "auditory STM" will be used in this report to represent "auditory and/or articulatory STM."

*S*s make different types of judgments with respect to successively presented pairs of visual stimuli: "Do the pairs have at least one item in common?" "Do they have both items in common?" "Are they identical?" (Do they have the same items in the same positions?) Two types of stimuli were used: dot patterns arranged to form letters and dot patterns randomly arranged. Some aspects of the results led *E*s to conclude that the letters, as well as the random patterns, were represented in memory in visual-like form. (The possibility that the letters were *also* represented auditorily was not ruled out.) The main evidence for this conclusion was several indications of strong position effects; for example, matches between items were more readily detected when those items occupied the same spatial positions in both target and probe displays than when they did not.

Snodgrass (1971) has also reported an experiment in which SAME–DIFFERENT judgments were made with respect to both digits and random visual patterns. The stimuli being compared were presented either simultaneously ("discrimination task") or at a temporal separation of a few seconds ("memory task"). In the latter case the procedure permitted the interval between stimuli to vary somewhat around either 2.5 or 4 sec. Snodgrass assumed that when the stimuli to be compared were presented simultaneously the comparisons would be largely visual and that long response times (RTs) would be indicative of difficult visual discriminations. The fact that positive correlations were found between memory and discrimination RTs for the same stimuli was taken as some evidence that both types of stimuli were encoded visually for the memory task. Further support for the conclusion that the digits in particular were visually encoded was seen in the fact that for both discrimination and memory tasks DIFFERENT RTs increased with an independent measure of visual similarity between the stimuli being compared.

The present study was designed to provide additional data on the question of visual versus auditory encoding of visual stimuli. The main assumption underlying the experiment was that if letter stimuli are stored in a different way than are random dot patterns, one might expect the different representations to decay at different rates. If they do, one should find that stimulus type (vis a vis auditory codability) and the amount of time the stimulus must be held in memory should interact in their effects on performance measures such as response time and error rate. In particular, we would expect these performance measures to be less dependent on the duration of the retention interval, within the range studied (0 to 8 sec.), with auditorily encoded stimuli than with stimuli for which auditory representations are not readily available.

It also seems reasonable to expect performance with these two types of stimuli to be differentially affected by changes in the exposure duration of the stimulus to be remembered. If it is the name of a stimulus that is stored rather than its visual configuration, performance should be independent of stimulus exposure duration, provided, of course, that the latter is great enough to permit the stimulus to be identified and its name retrieved. When the visual pattern itself must be retained, we might expect that, within limits, the longer the stimulus duration the more adequate will be the memory representation that is formed (Shaffer & Shiffrin, 1972). This conjecture would be supported by an interaction between stimulus type and exposure duration.

Finally, there is the problem of distinguishing between the effects of auditory codability and those of such factors as familiarity and discriminability with which they may be confounded. No attempt will be made to unconfound auditory codability and familiarity; it will simply be assumed that these variables are thoroughly confounded in general, the more familiar the visual pattern the more likely it is to have a name. There is no necessity, however, for visual patterns with names to be more or less discriminable than those without them, although one might guess that they would tend to be more so. To help distinguish between the effects of auditory codability and discriminability in the present experiment, pairs of random patterns representing two levels of discriminability were used. By comparing performance obtained with these two sets of stimuli against that obtained with letters, it should be possible to determine whether performance differences associated with the letter-versus-random-pattern distinction may be attributed solely to the greater (or lesser) discriminability of the letters, or must be accounted for in terms of some other type of difference.

METHOD

Task

The *S*'s task was to compare two successively presented visual patterns and to report as quickly as possible, by pressing one of two response keys, whether the patterns were the same or different. The first stimulus was always masked before the second was displayed.

Independent Variables

There were four independent variables: (*a*) stimulus type, (*b*) stimulus exposure duration, (*c*) retention interval, and (*d*) type of response required (positive or negative).

Stimulus type.—Stimulus patterns were formed from subsets of a 7 × 5 dot matrix (see Nickerson & Pew, 1972, for scale drawings of similar patterns used in another experiment). A dot matrix measured $\frac{6}{16}$ in. (height) × $\frac{5}{16}$ in. and was viewed from a distance of about 2 ft. Under one condition the stimulus patterns formed block capital English letters; under two other conditions they were whatever configurations occurred when dots were placed at random in 12 of the 35 possible positions. The two latter conditions differed with respect to the magnitude of the difference between differing pairs. In the "easy random" condition the two stimuli of any pair either were the same or differed with respect to the position of 11 of their 12 dots; in the "difficult random" condition the members of a pair either were the same or differed with respect to the position of a single dot.

Stimulus exposure duration.—The first stimulus of each pair was exposed for .25, 1, or 4 sec. The exposure of the first stimulus was terminated by the superimposition of a 7 × 5 dot pattern on the display area. This masking pattern remained on until *S* made his response. The second stimulus appeared directly under the first with a $\frac{5}{16}$-in. separation distance between the stimuli.

Retention interval.—The interval between the onset of the mask and that of the second stimulus was varied among six durations: 0, .25, .5, 1, 2, and 8 sec.

Type of response required.—On exactly half of the trials of every session the correct response was SAME. The experiment was counterbalanced with respect to which hand was used to make this response.

Design

Stimulus type and stimulus exposure duration were treated as between-*S*s variables. Both the duration of the retention interval and the type of response required were varied randomly from trial to trial, within the constraint that in every run of 60 trials, each of the six durations occurred 10 times, 5 times with positive (SAME) stimuli and 5 times with negative.

Procedure

Stimuli were presented on a cathode-ray tube display under the control of a Digital Equipment Corporation PDP/8 computer. Figure 1 shows the layout of the display and the temporal structure of each trial. At the beginning of each trial there was displayed for 1 sec. at Position 1 the duration of the retention interval for that trial. This was done to minimize temporal uncertainty, and thereby, hopefully, to preclude effects that could be attributed

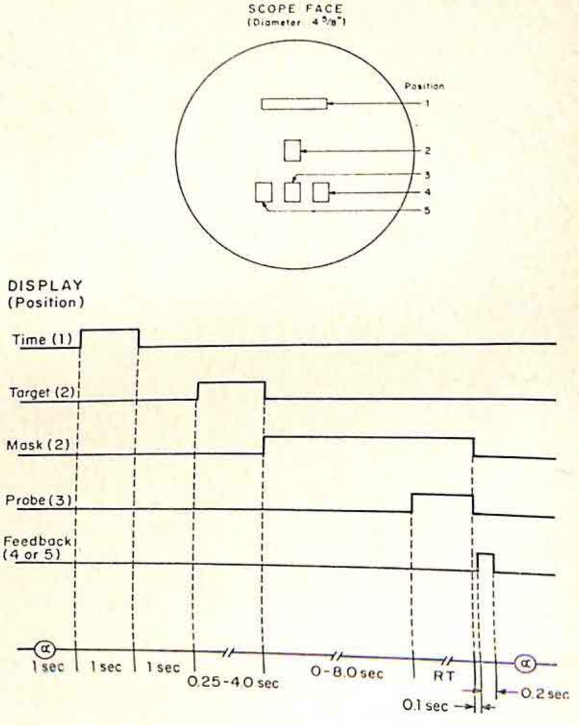


FIG. 1. The layout of the visual display and the temporal structure of a trial. ("Target" refers to the first of the two stimuli being compared, and "probe" refers to the second.)

to the possible selection of inappropriate rehearsal strategies that might be used when *S* does not know how long he must retain information. One second following the offset of this display, the first stimulus of a pair was presented at Position 2. The duration of this display was either .25, 1, or 4 sec., depending on the condition. At the end of this interval a masking stimulus was presented at Position 2. Following a delay of from 0 to 8 sec., during which time the masking stimulus remained on, the second stimulus of the pair being compared was presented at Position 3. This stimulus was displayed, along with the masking stimulus, until a response was made. Shortly (.1 sec.) after the response was made either an "S" or a "D" appeared in either Position 4 or Position 5. The letter indicated the response that *should* have been made, irrespective of which one had been; and it appeared on the side of the display corresponding to the hand to which that response had been assigned. The feedback display remained on for .2 sec. One second following its offset the next trial was initiated with the onset of the display of the next retention interval at Position 1.

Each run had 66 trials. The first 6 (in which each of the six retention intervals occurred once, and half of the stimuli were SAME) were considered practice, and data from those trials were not retained.

Each *S* participated during 2 days. On the first day the procedure was explained in detail, and at least three practice runs were given. Four and sometimes five practice runs were given on the more difficult conditions. On the second day each *S* had five runs, and these data are reported.

The *S*s rested between runs as they wished. Usually they ran in pairs, one resting while the other ran. The *S*s were permitted to see their mean SAME and mean DIFFERENT RTs as well as the number of correct responses of both types following each run.

Subjects

The *S*s were 18 high school students who were paid for their participation. Two *S*s were assigned (on an as-arrived basis) to each of the nine combinations of the two between-*S*s variables.

RESULTS

Figure 2 shows RT (for correct responses) as a function of the duration of the retention interval for each combination of stimulus type and exposure duration. Each point in Panels a, b, and c is the mean of four medians, a median SAME and a median DIFFERENT RT for each of two *S*s. Each median was based on a maximum of 25 RTs, the actual number depending on the number of errors associated with that set of scores. In Panel d, data have been pooled over exposure duration; the dashed line shows them pooled over both exposure duration and stimulus type.

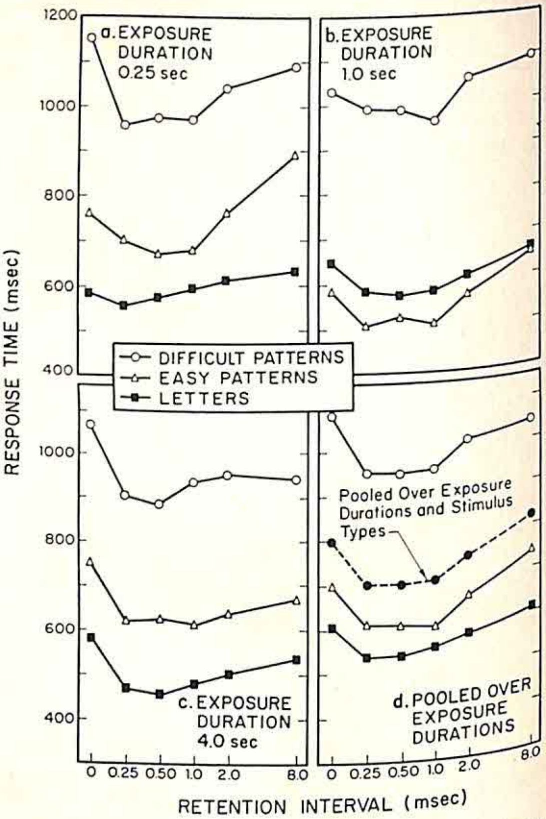


FIG. 2. Response time as a function of the duration of the retention interval for the various experimental conditions. (Panel d represents pooled data.)

An analysis of variance on the median correct RTs on which Fig. 2 is based showed the main effects of stimulus type, retention interval, and response type to be significant at well beyond the .01 level. The effect of exposure duration was not significant. The only interactions that were significant at a .05 level or better were Exposure Duration × Retention Interval and Exposure Duration × Response Type. The Stimulus Type × Response Type interaction was nearly significant.

The effects of stimulus type and retention interval are clear from the figure. The DIFFICULT random patterns took much longer to compare than did either the "easy" random patterns or the letters.[4]

---

[4] The "easy patterns" curve in Panel b may be spuriously low; one of the two *S*s represented by that curve appeared to be unusually fast.

For all combinations of exposure duration and stimulus type, the relationship between RT and retention interval was characterized by an initial dip and then a rise, minimum RTs being obtained with intervals of intermediate duration. The similarity of these curves is consistent with the lack of an interaction between stimulus type and interval.

It appears to be the case that the slope of the linear component of the curves decreases somewhat as exposure duration increases, which accounts for the significant Exposure Duration × Retention Interval interaction. If Panels a and c of Fig. 2 were overlaid, it would be seen that exposure duration had little effect on RT at the shortest retention interval, but a large effect at the longest ones, the longer exposure duration producing the shorter RTs.

The effects associated with response type are not seen in Fig. 2 inasmuch as the curves shown there represent positive and negative responses combined. In general, negative RTs were longer than positives—by about 38 msec. on the average (760 msec. vs. 722). The Exposure Duration × Response Type interaction indicates that the magnitude of the difference between positives and negatives was not constant over exposure durations, although the difference was in the same direction in all cases. The nearly significant Stimulus Type × Response Type interaction reflects the fact that the difference between negative and positive RTs was greater (about 50 msec.) for the letters and easy random patterns than for difficult random patterns (about 10 msec.).

Figure 3 shows percent errors for each of the conditions represented in Fig. 2. An analysis of variance was performed using as data four scores associated with each point in Panels a, b, and c. The four scores were percent incorrect SAMES and percent incorrect DIFFERENTS for two *S*s. The results of the analysis are tenuous inasmuch as homogeneity of variance over stimulus types is not a valid assumption; there is an obvious floor effect for letter stimuli. The results of the analysis are reported here primarily for the sake of calling attention to specific relationships that can be observed in Fig. 3.

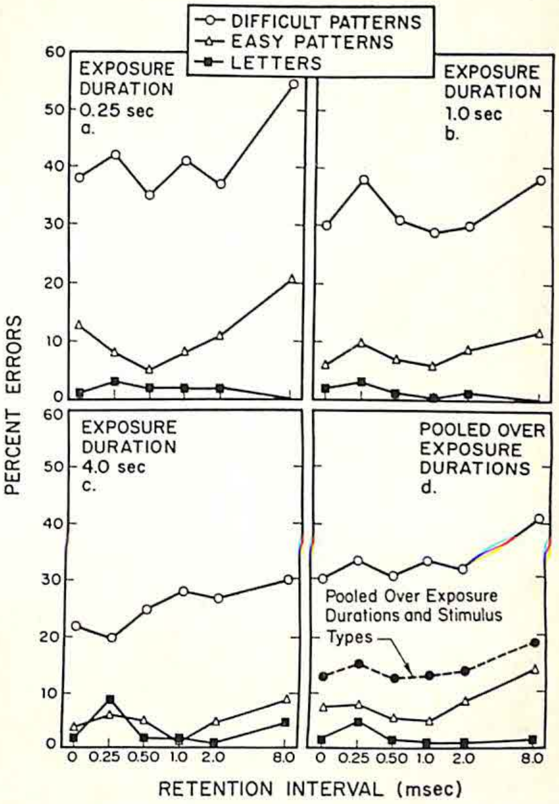


FIG. 3. Error rate as a function of the duration of the retention interval for the various experimental conditions. (Panel d represents pooled data.)

Each of the main effects was significant at the .01 level or better, as were three interactions: Exposure Duration × Stimulus Type × Response Type, Exposure Duration × Interval × Response Type, and the four-way interaction involving all factors. The Exposure Duration × Stimulus Type interaction was significant at the .05 level.

The effects of stimulus type and of exposure duration are apparent in the figure. So is that of the Exposure Duration × Stimulus Type interaction. The main effect of retention interval is less obvious, given the relative flatness of the dashed curve in Panel d. The effects involving response type are not visible in the figure because the two types of errors are not differentiated. In general, false SAMES (the response SAME when the stimuli were different) were more frequent (16.0%) than

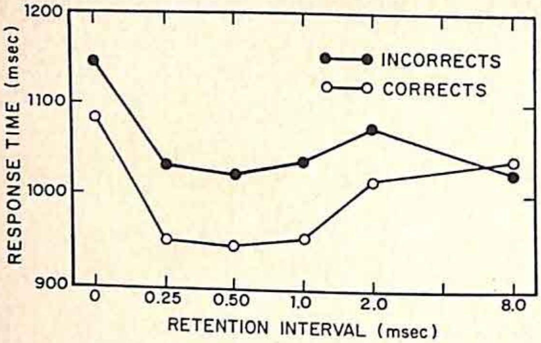


FIG. 4. Response time as a function of the duration of the retention interval for both correct and incorrect responses. (Only data obtained with the difficult random patterns are represented.)

were false DIFFERENTS (12.8%).[5] The interactions reflect the fact that this split was not consistent for different combinations of stimulus type, exposure duration, and interval.

Median RTs were obtained for the incorrect responses of each *S* for each of the exposure duration retention interval combinations for the difficult random patterns. These medians were an average of 57 msec. longer than the corresponding medians for correct responses. The relationships between these medians and exposure duration or retention interval were not appreciably different from those observed with correct responses. Figure 4 shows the mean median RTs for correct and incorrect responses as functions of interval with data pooled over exposure durations. (The curve representing correct responses is the same as the top curve in Fig. 2d.) The curves are nearly identical in shape except for the right-most points.

Error RTs were not considered for easy random patterns or for letters because of the low frequencies of occurrence in those cases. Several of the cells representing particular combinations of exposure duration, interval, error type, and *S* had no responses; and several others had only one or two.

DISCUSSION

Visual versus Auditory Memory Representation

The results provide little evidence that the memory representations for the letters and random patterns used in this experiment decayed at different rates. Stimulus type and retention interval did not interact in their effects on either response times or error rates. The data do suggest that the different types of stimuli may have required different amounts of time to be encoded for memory storage. This is seen in the fact that the degree to which error rates were decreased by increasing exposure duration depended on the type of stimulus involved. In particular, the effect was greatest in the case of the difficult random patterns. This result must be interpreted cautiously, however, inasmuch as error rates with the letters were minimal even with the shortest stimulus exposure duration, and hence could not be much reduced as the exposure duration was increased. On balance, there is little in the results that would compel us to conclude that the random dot patterns were encoded and represented in memory in a qualitatively different way than were the letters. The easy random patterns produced results that were more similar to those obtained with letters than to those obtained with the difficult random patterns. On the whole, the data are consistent with the idea that the main difference among the three sets of stimuli was their relative discriminability. In particular, they suggest that although the discrimination that was required in the case of the difficult random patterns was very difficult indeed, relative to that required in the case of letters, the basis on which it was made was not qualitatively different.

This conclusion is difficult to reconcile with the prevailing view that nameable visual stimuli are represented in STM in an auditory or articulatory code. Three possible explanations are offered for the apparent failure to obtain evidence of auditory encoding in the present case. (*a*) The assumption that visual and auditory representations of the stimuli used in this experiment would decay at different rates may be invalid. If it is, then the results could not be taken as evidence that both types of stimuli were stored in the same form. (*b*) The results may depend critically on the fact that *S* had to retain only a single item at a time. It is generally held that at least one of the reasons for encoding nameable visual stimuli auditorily is to facilitate rehearsal, the assumption being that *S* can rehearse auditory representations better than visual. Maybe the assumption is valid only for *sequences* of stimuli. Given the task of remembering a single visual stimulus for a short period of time, perhaps it is as easy to maintain the visual representation as to recode it into its auditory associate (Posner, Boies, Eichelman, & Taylor, 1969). (*c*) Finally, it may be that the data obtained in this experiment were simply too noisy to permit the detection of differences that exist. There is in fact some suggestion in the data, although it is not statistically substantiated, of differential decay rates. For the .25-sec. exposure duration condition, the curves (both RT and error rate) representing letters and easy random patterns appear to be diverging at the longer retention intervals. Extending the range of intervals might verify the effect. Assuming for the moment that the suggested effect is real, the following question arises: Why should stimulus exposure duration affect the rate of memory decay? It is easy to see why a brief exposure would result in an inadequate memory representation to begin with, but given that the representation is good enough to permit fast and accurate responses at short intervals (e.g., 500 msec.), why should its decay be accelerated thereafter?

---

[5] These percentages, like all others reported, are percentages of the number of events of the specified type that could have occurred. For example, 16% false SAMES means that of those stimuli for which the correct response was DIFFERENT 16% were responded to by SAME.

The Relationship between RT and Retention Interval

The shapes of the curves relating RT to retention interval deserve some comment. In particular, it is interesting that the curves do not uniformly increase or decrease monotonically. The typical shape is an initial decrease followed by a rise. In the foregoing discussion we have implicitly attributed the increase in RT with the longer interval durations to deterioration of the memory representation, or forgetting. However, there is another factor whose effect cannot be ruled out, and that is time uncertainty. We usually think of *S*'s uncertainty about the time of arrival of a signal as having two sources: the objective uncertainty due to the trial-to-trial variability of the signal's arrival time, and *S*'s imperfect timekeeping ability (Klemmer, 1956, 1957). In this experiment the first source of uncertainty was eliminated by telling *S* on each trial exactly what the duration of the retention interval would be. There remained, however, his imperfect timekeeping ability, the effect of which we would expect to increase with the duration of the interval that must be judged. This could account, at least in part, for the increase in RT with increasing interval durations. It would be of interest to do an experiment to attempt to determine how much of the increase in RT with retention interval may reasonably be attributed to this factor. To do this one would have to provide *S* with a timekeeping cue to compensate for his imperfect internal clock. For example, one might provide on each trial some analog representation of time, such as a line whose length decreased at a constant rate and which disappeared at the precise instant that the test pattern was displayed. Such a time cue would be unlikely to be effective with the shorter retention intervals, but might be helpful in the case of the longer ones.

The initial dip in the RT curves may also have little to do with memory processes. To be sure, several *E*s (Nickerson, 1967b; Posner et al., 1969; Snodgrass, 1971) have found that comparison RTs are often longer when the items being compared are presented simultaneously than when the onset of one precedes that of the other, even by a fraction of a second. In these experiments the difference between RTs obtained with simultaneous and sequential presentations can be accounted for by assuming that in the latter case *S* can assimilate or encode one of the stimuli before the other one is presented, thus getting a headstart on the processing, which is not possible when the items are presented sequentially. However, in the present experiment the stimuli were presented sequentially even when the interitem interval was 0 msec. So this explanation does not apply.

There are two plausible explanations for the relatively long RTs obtained with the 0-retention interval. It may be that *S* waited until the first stimulus was masked before changing his point of fixation from Position 1 to Position 2 (see Fig. 1), in which case RT would be long when the interval was 0 msec. because it would include part of the time to make the eye movement. Alternatively, one might attribute the long RTs to an inhibiting effect of the onset of the mask. The literature on "psychological refractory phase" makes it clear that a stimulus such as that used as a mask in this experiment can delay the response to a simultaneous or closely following stimulus under a variety of conditions (Davis, 1959, 1962; Nickerson, 1967a).

SAME versus DIFFERENT RTs

The relationship between SAME and DIFFERENT RTs has been the subject of considerable attention of late. The most typical finding seems to be that SAME RTs tend to be shorter than DIFFERENTS, although in some cases the reverse finding has been reported. Among the conjectures that have been put forth to account for these differing results is one that holds that SAME RTs will be shorter than DIFFERENTS when the stimuli are represented in memory in a verbal code, as letters, digits or words might be; and the reverse will be the case when the memory representation has more the character of a sensory image, as, for example, pure tones differing in frequency or amplitude, or lines differing in length (Bindra, Donderi, & Nishisato, 1958). In the present experiment, SAME RTs were significantly shorter than DIFFERENTS, on the average. However, the size of the difference varied as a function of stimulus type. In particular, it was relatively large for the letters, somewhat smaller for the easy random patterns, and small, almost to the point of vanishing, for the difficult patterns. The results obtained with the letters, therefore, were consistent with the verbal encoding conjecture. The results obtained with the random patterns, however, are problematic for this notion. The lack of a sizable difference either way with the difficult patterns might be discounted as indicative of nothing in particular, given the huge error rates obtained with these stimuli. However, the fact that SAME RTs were considerably shorter than DIFFERENTS with the easy random patterns is not consistent with the verbal encoding notion.

Finally, note that the discriminability variable affected not only DIFFERENT RTs but SAMES as well, and this is so, in spite of the fact that matching pairs of stimuli in the difficult random condition were essentially equivalent to matching pairs in the easy random condition. The only difference between these two conditions was the magnitude of the difference, when the stimuli differed. This substantiates the argument that has been made in other places that how long it takes to decide that two stimuli are the same must depend on the magnitudes of the differences that *S* may expect to encounter (Nickerson, in press).

REFERENCES

BINDRA, D., DONDERI, D. C., & NISHISATO, S. Decision latencies of "same" and "different" judgments. *Perception & Psychophysics*, 1968, **3**, 121–130.

CONRAD, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75–84.

DAVIS, R. The role of "attention" in the psychological refractory period. *Quarterly Journal of Experimental Psychology*, 1959, **11**, 211–220.

DAVIS, R. Choice reaction times and the theory of intermittency in human performance. *Quarterly Journal of Experimental Psychology*, 1962, **14**, 157–166.

KLEMMER, E. T. Time uncertainty in simple reaction time. *Journal of Experimental Psychology*, 1956, **51**, 179–184.

KLEMMER, E. T. Simple reaction time as a function of time uncertainty. *Journal of Experimental Psychology*, 1957, **54**, 195–200.

NEISSER, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

NICKERSON, R. S. Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology*, 1965, **19**, 155–160.

NICKERSON, R. S. Psychological refractory phase and the functional significance of signals. *Journal of Experimental Psychology*, 1967, **73**, 303–312.

NICKERSON, R. S. "Same"–"different" response times with multiattribute stimulus differences. *Perceptual and Motor Skills*, 1967, **24**, 543–554.

NICKERSON, R. S. The use of binary-classification tasks in the study of human information processing: A tutorial review. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press, in press.

NICKERSON, R. S., & PEW, R. W. *Visual pattern matching: An investigation of some effects of decision task, auditory codability and spatial correspondence.* (BBN Rep. No. 2392) Cambridge, Mass.: Bolt Beranek and Newman, Inc., 1972.

NOTON, D., & STARK, L. Eye movements and visual perception. *Scientific American*, 1971, **224**(6), 34–43.

POSNER, M. I., BOIES, S. J., EICHELMAN, W. H., & TAYLOR, R. L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology*, 1969, **79**(1, Pt. 2).

SHAFFER, W. D., & SHIFFRIN, R. M. Rehearsal and storage of visual information. *Journal of Experimental Psychology*, 1972, **92**, 292–296.

SHEPARD, R. N. Recognition memory for words, sentences and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 156–163.

SNODGRASS, J. G. Matching patterns vs. matching digits: The effect of memory dependence and complexity on "same"–"different" reaction times. Paper presented at the meeting of the Psychonomic Society, St. Louis, November 1971.

SPERLING, G., & SPEELMAN, R. G. Acoustic similarity and auditory short-term memory: Experiments and a model. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

WICKELGREN, W. A. Short-term recognition memory for single letters and phonemic similarity of retroactive interference. *Quarterly Journal of Experimental Psychology*, 1966, **18**, 55–62.

(Received March 3, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 437–444

TEMPORAL ASPECTS OF SIMPLE MULTIPLICATION AND COMPARISON[1]

JOHN M. PARKMAN[2]

Carnegie-Mellon University

For six sessions, six college students were shown series of simple multiplication problems of the form $p \times q = r$ and were asked to respond rapidly whether the solution given with each problem (r) was correct or not. On half of the trials, the presented solution equaled the true product and, on the remaining trials, the presented solution differed from the true product in a systematic way. Additive multiplication-stage and comparison-stage reaction time effects were found for correct responses. For the multiplication stage, latencies were found to increase monotonically as a function of the sum of p and q and as a function of $\min(p,q)$, excluding problems where p or q equaled zero. For the comparison stage, negative responses had longer latencies than did positive responses. These results were found to be very similar to those from an earlier study of simple addition. Some theoretical implications of this similarity were explored.

For most adults, the multiplication of two single-digit numbers is a highly overlearned task performed quite rapidly and with little conscious effort. For example, if *S* is asked to multiply 7 times 5 and then to retrospect on what he did, he will usually report that he simply knew or remembered the correct answer, or equivalently, that somehow he simply became aware of 35. Questioned further on how he arrived at the solution, he will generally plead ignorance, except to state that of course he learned the answer much earlier and, hence, all he did was to recollect a previously learned fact.

Certainly this explanation is plausible enough, given our current state of knowledge. Most adults have had a long history of dealing with multiplication facts like "seven times five" and have had ample opportunity to form long-term representations of the correct answers. But beyond this very little is known. Indeed, even this "information retrieval hypothesis" of simple mental multiplication is based on very little direct empirical evidence and gains most of its plausibility from casual observation and introspection.

One approach to the study of simple mental multiplication and its underlying processes is to investigate reaction time differences among the basic multiplication problems. Such differences may well provide clues to the informational processes involved in this and related computational tasks. The present study investigated reaction times associated with the 100 basic multiplication problems formed from all possible pair-wise combinations of the single digits 0–9. These reaction times were measured indirectly using a procedure similar to that used in a study of simple addition (Parkman & Groen, 1971). Thus, in the present study, problems were presented in the form $p \times q = r$ and *S*s were asked to decide whether or not the displayed answer (r) was correct, and the total reaction time was assumed to be the sum of the duration of component processes, one of which was the multiplication process.

There have been very few reaction time studies of mental multiplication in adults, and most of these have investigated solution times to multidigit problems in which considerable time is needed for carry operations, storing of intermediate results, etc. For example, Dansereau and Gregg (1966) found that the solution times to problems involving two-, three-, or four-digit numbers multiplied by a two-digit number are well described by a "difficulty factor" based on the number of single-digit operations involved in each problem (e.g., "multiply," "add," "hold," "carry"). Along these lines, Dansereau (1969) constructed a computer simulation model of multidigit mental multiplication performance. As interesting as these studies are in terms of the control processes used in a complex cognitive task, they provide no information concerning the basic multiplication operation. Dansereau, for example, assumed that this operation is composed of a table look-up procedure but did not attempt further specification of the processes involved.

---

[1] This research was supported in part by a grant from the Public Health Service, National Institute of Mental Health, Research Grant MH-07722. The author wishes to thank Patrick Cavanagh, William Chase, and Guy Groen for their critical comments.

[2] Requests for reprints should be sent to John M. Parkman, Department of Psychology, State University of New York at Buffalo, 4230 Ridge Lea Road, Buffalo, New York 14226.

Thomas (1963), on the other hand, has dealt with the multiplication operation more directly. He assumed that the mental processes underlying many forms of simple calculation behave as an information-processing system for which the time needed to perform a given arithmetic operation is strictly proportional to the amount of information that must be handled by the system. Given this premise and some notions concerning how the structural components of a given problem relate to one another in terms of total information, Thomas derived the prediction that reaction times to simple multiplication problems of the form $p \times q = r$ should be a linear function of $\log(p + q + r)$. However, Thomas did not specify how such a function might be related to psychological processes.

A fundamental question with regard to mental multiplication is whether *S* forms the product of two single-digit numbers by use of some memory retrieval procedure or whether such a product is generated by a computational algorithm. Reaction time data alone cannot settle this question but can, at least, provide conditions that must be met by any model of the multiplication process.

An analogous situation exists for simple mental addition, for which some systematic reaction time data have recently been collected. Specifically, Parkman and Groen (1971) found that reaction times to problems of the form $p + q = r$ are in general a linear function of the sum of the two addends and to about the same extent a linear function of the minimum addend. As they point out, this latter finding is compatible with one of a class of counting models for simple addition proposed earlier by Suppes and Groen (1967) and extensively discussed by Groen and Parkman (1972). Very briefly, the model in question consists of two basic operations: the setting, or initialization, of a counter and the successive incrementation of the counter until the desired result is obtained. The model performs these operations in the following sequence: first the counter is set to the value of the larger addend, and then the counter is successively incremented by one *n* times, where *n* is the value of the smaller addend. It is assumed that the time required to set the counter is independent of the value to which the counter is set and that the time needed to increment the counter is independent of the number of times the counter is incremented. Hence, the model predicts that reaction times for simple addition problems will be a linear function of the minimum addend with the slope of this function estimating the duration of each incrementation cycle.

Though it does appear plausible that adult *S*s perform simple addition by means of a high-speed counting operation, as Parkman and Groen (1971) suggest, it is nevertheless true that their data are also compatible with some types of memory retrieval models. For example, the counting algorithm they propose could easily be modified to calculate an index, rather than a sum, with the index then being used for a memory retrieval operation. This latter formulation has the advantage that it can easily be generalized to encompass simple mental multiplication, while the original model can do so only with the rather strong assumption that the model's counting procedure can operate by increments of arbitrary size rather than the basic counting by ones which Suppes and Groen (1967) assumed.

From the above, it can be argued that if reaction time data for simple multiplication have the same general characteristics as those found for simple addition, then perhaps the counting model for simple addition proposed earlier should be reconsidered. Furthermore, the possibility would then exist that addition, multiplication, and perhaps other forms of simple calculation are all based on common processing mechanisms operating in such a way as to generate similar reaction time effects. On the other hand, if close similarities are not found between simple mental addition and multiplication, then it may well be that performance in the two tasks is based on quite different processes. Hence, one goal of the present study was to examine general similarities and differences between simple multiplication reaction times and the reaction time data collected by Parkman and Groen (1971) for simple addition.

METHOD

Subjects.—Six students from Carnegie-Mellon University, five males and one female, served six sessions each and were paid for participation. Three *S*s had participated in previous reaction time experiments. All *S*s were right handed.

Material.—The multiplication problems were of the form $p \times q = r$, where p and q were the single, nonnegative integers 0–9. All possible pairings of these single digits were used on the left side of the equal sign. Each of the resulting 100 multiplication problems was paired with its correct product (positive instance) and with numbers differing from its correct product (negative instances). Negative instances for a given problem ($p \times q$) were formed by either incrementing or decrementing either p or q by one and then calculating the resulting product (r). These resulting products were in turn constrained to nonnegative values. (When a negative instance was needed for the problem 0×0, r was set equal to five.)

Apparatus.—A DDP-116 digital computer controlled presentation of problems via an on-line video monitor. The problems appeared centered on the monitor's screen, filling an imaginary rectangle .89 cm. in height and 3.01 cm. in width. The *S*s viewed the screen from an eye distance of approximately 63.5 cm. which subtended a visual angle of 2°43′. Responses were made by pressing one of two microswitches located approximately 50 cm. in front of the screen. The positions of the positive and negative response switches were counterbalanced across *S*s. Latencies were timed and recorded by the computer to the nearest .001 sec. from the time the problem was sent to the monitor until switch contact was made. Images appeared on the monitor within approximately .008 sec. after the appropriate commands had been issued and latency timing initiated.

Design.—The *S*s were run one session per day for 6 days. For each *S*, the first session was conducted 3 days before the second session, and Sessions 2–6 were conducted on successive days. Each session was divided into four blocks of 100 trials each. Across each session, *S* saw the 100 different multiplication problems paired twice with the respective correct solution and twice with incorrect solutions. On the latter two pairings, one presented solution was greater than the true product while the other presented solution was smaller than the product, except when the multiplication problem involved a zero term, in which case both pairings were larger than the true product. Hence, during each session, 50% of the problems were paired with the correct solution. Furthermore, within each block, *S* saw 50 problems with correct solutions and 50 problems with incorrect solutions. Excluding problems with zero terms, 50% of the negative-instance problems within each block had presented solutions which were larger than the true product and 50% had presented solutions which were smaller than the true product. The decision to adjust either p or q in generating negative instances was randomly made by the computer just prior to the presentation of each negative-instance problem. Within blocks, trials were randomized within the following sequential trial constraints: (*a*) No single digit could appear as p or q in two successive problems, e.g., $5 \times 7 = 35$ could not follow $7 \times 3 = 21$. (*b*) No two problems having the same product could appear on successive trials, whether or not the product in question actually appeared with the respective problems, e.g., $6 \times 4 = 18$ could not follow $3 \times 8 = 32$. These sequential constraints were added in order to minimize possible repetition effects.

Procedure.—The *S*s were run individually, with each session lasting approximately 20 min. The task was to answer whether or not the solution presented with a given multiplication problem was correct by pressing the appropriate switch and to answer "as quickly and accurately as possible."

A 15-sec. interblock rest period was used. During this period, the word "REST" appeared on the screen. Within blocks, trials were conducted in the following sequence: (*a*) The word "READY" appeared on the monitor screen for .5 sec. within the imaginary rectangle mentioned above. (*b*) The screen was blank for .5 sec. (*c*) The problem-solution pair appeared until *S* responded. (*d*) The screen remained blank 1.0 sec. prior to the presentation of the next problem. However, if *S* responded incorrectly, the above sequence was modified in the following way, beginning with *S*'s response: (*i*) The screen remained blank for .5 sec. (*ii*) The word "WRONG" appeared for 1.0 sec. where "READY" had appeared previously. (*iii*) The screen remained blank 1.0 sec. prior to the presentation of the next problem.

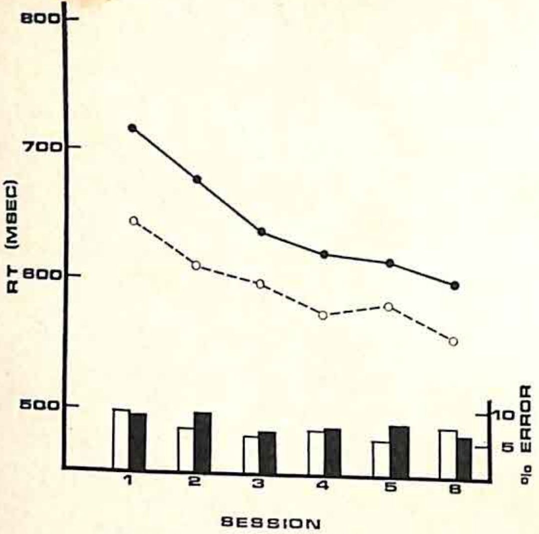


FIG. 1. Mean RT as a function of session. (Filled circles and bars refer to negative instances, open circles and bars refer to positive instances.)

RESULTS

Figure 1 presents the mean latencies for correct responses, pooled over *S*s, plotted separately for positive and negative instances as a function of session. Also presented in Fig. 1 are the mean error rates for both positive instances (false negative responses) and negative instances (false positive responses). The overall mean error rate was .070. Individual *S* error rates ranged from .063 to .095.

The data reported in the remainder of this section are based solely on Sessions 2–6 unless otherwise noted; i.e., Session 1 is treated as a practice session for all *S*s. In addition, all reaction time data are based only on correct responses.

Each *S* required more time, on the average, to respond to negative instances than to positive instances, the range for *S*s, pooled across sessions, being from 28 to 56 msec. Table 1 shows the mean reaction times, pooled over *S*s and sessions, as a function of positive and negative instances. (The *SE*s are based on the means for the six *S*s). The 45.8-msec. mean difference between positive and negative instances is significant, using a *t* test for dependent groups, $t\ (5) = 7.09$, $p < .001$. However, for negative instances, the 5.9-msec. mean difference between problems whose displayed solutions were numerically larger than the correct product and problems whose solutions were smaller is not significant, $t\ (5) = 1.73$, $p < .20$.

For purposes of further analysis, the individual-problem reaction time data were partitioned into three sets based upon the digits ($p \times q$) composing each problem. The three sets were: (*a*) all problems for which p and q were not equal and for which neither p nor q was equal to zero (Min 1–8, non-tie problems), (*b*) problems for which p and q were not equal and either p or q were zero (Min 0, non-tie problems), and (*c*) all problems with p equal to q (ties). The ties were removed from the other problems in order to facilitate comparison of results with the Parkman and Groen (1971) addition results; in their study, ties were found to have substantially faster reaction times than other problems and were subjected to separate analysis. The Min 0, non-tie problems were treated separately in the present study due to their unexpectedly long reaction times. In each of the three sets, the mean difference between positive- and negative-instance data was approximately equal to the value reported above and was always significant ($p < .01$).

Min 1–8, non-tie problems.—The 72 basic problems in this set were studied as a function of each of the five structural variables investigated by Parkman and Groen (1971), viz., the Min $[\min(p,q)]$, the Max $[\max(p,q)]$, the Split $|p - q|$, the Sum $(p + q)$, and Thomas's statistic $[\log(p + q + r)]$, with p, q, and r defined as above. Linear regressions were performed between each of these structural variables and the mean reaction times for

TABLE 1

MEAN REACTION TIMES FOR POSITIVE AND NEGATIVE INSTANCES

Measure	Instance type			
	Negative			Positive
	Larger	Smaller	Average	
M	636.2	630.3	633.3	587.5
SE	55.8	49.1	52.5	54.1

Note.—Reaction times are measured in milliseconds.

TABLE 2

INTERCORRELATION MATRIX OF STRUCTURAL VARIABLES FOR 72 BASIC MULTIPLICATION PROBLEMS

Variable	1	2	3	4
1. Min				
2. Max	.500			
3. Split	−.500	.500		
4. Sum	.866	.866	.000	
5. Thomas's statistic	.898	.788	.111	.973

the various problems. Before turning to the results, however, it should be noted that given the design used in the present study and the set of 72 problems under investigation, the structural variables are a priori intercorrelated to varying degrees. Table 2 gives these intercorrelations. The rather high correlation between some pairs of structural variables simply means that if the two variables in such pairs are found to be significantly and about equally related to reaction time, further research may be needed to determine whether or not one of the variables' significant relationship may be due to spurious correlation.

Table 3 shows the slope estimates and R^2 values for each of the five structural variables based on the 72 positive-instance means. Table 4 displays results from an identical analysis using the 72 negative-instance means. As can be seen in Table 3 and Table 4, the Min, the Sum, and Thomas's statistic all have strong relationships with the reaction time data. It can also be seen that these relationships are approximately equal for both positive-instance and negative-instance problems. The overall mean reaction time for these problems, pooling over positive and negative instances, was 604.5. msec.

TABLE 3

COMPARISON OF R^2 AND SLOPE ESTIMATE VALUES FOR STRUCTURAL VARIABLES (POSITIVE INSTANCES)

Variable	R^2	Slope estimate
Min	.627	18.7 msec/unit
Max	.261	12.1 msec/unit
Split	.079	−6.6 msec/unit
Sum	.567	10.3 msec/unit
Thomas's statistic	.683	126.3 msec/unit

TABLE 4

COMPARISON OF R^2 AND SLOPE ESTIMATE VALUES FOR STRUCTURAL VARIABLES (NEGATIVE INSTANCES)

Variable	R^2	Slope estimate
Min	.633	15.3 msec/unit
Max	.256	9.7 msec/unit
Split	.084	−5.6 msec/unit
Sum	.565	8.4 msec/unit
Thomas's statistic	.675	102.2 msec/unit

Comparing the results in Table 3 and Table 4 with the corresponding data from the Parkman and Groen (1971) study of simple addition, one finds a rather high degree of similarity in the pattern of R^2 values. However, there are two minor differences between these sets of data. First, problems with a Min of zero were not included in the analyses performed for Table 3 and Table 4 of the present study but were included in the corresponding analysis performed for simple addition. The reason for this action will be discussed below. Second, multiplication problems with a Min of one had lower mean latencies, relative to problems with higher Mins, than was the case in addition.

This Min-of-one result can be seen in Fig. 2, where individual problem mean reaction times have been pooled and plotted as a function of the Min, with positive- and negative-instance data plotted separately. One consequence of these "fast" reaction times for Min-of-one problems in multiplication is that the present set of data are fit by the Thomas statistic better, relative to the Min and Sum, than was the case for addition. (This result is especially enhanced, given the current experimental design, because the smaller the Min the more individual problems having that Min. Hence, as the analyses performed for Tables 3 and 4 are based on individual-problem means, the R^2 values are heavily influenced by problems with low Mins.)

The problems in this set were also investigated for possible order effects, i.e., for problems of the form $p \times q$ to have reaction times differing from their symmetric pairs,

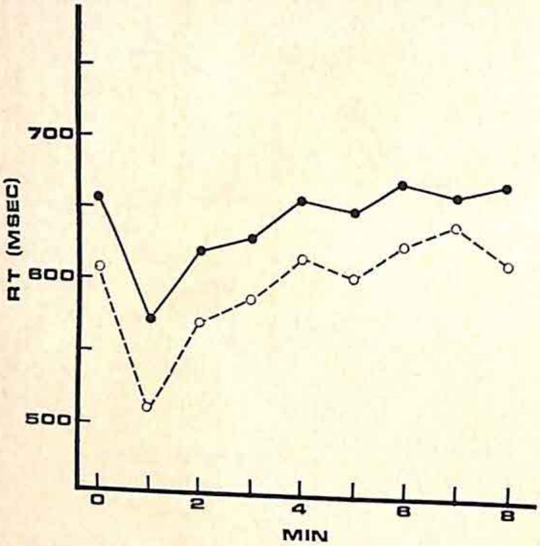


FIG. 2. Mean RT for non-tie problems as a function of the Min. (Filled circles refer to negative instances, open circles refer to positive instances.)

$q \times p$. Positive-instance problems were found to average 3.3 msec. faster with the Max in the left position, but negative-instance problems were 1.3 msec. slower in this configuration. Neither result approached significance, $t\ (35) = .158$, and $t\ (35) = .085$, both $ps > .50$, respectively.

Min 0, non-tie problems.—Parkman and Groen (1971) found that non-tie problems with a Min of zero had faster associated latencies than other non-tie problems in keeping with the general linear, increasing function they found between reaction time and size of Min. However, as can be seen in Fig. 2, this result was clearly not the case in the present study, as multiplication problems with a Min of zero had longer latencies than would have been true given such a relationship with the Min. Since the latencies for these problems were so much out of accord with the trend for the other non-tie problems, they were removed from the analysis to enable a clearer determination of the processing involved in non-tie problems.

The overall mean for this set of problems, pooling over positive and negative instances, was 631.8 msec. As with the other non-ties, there was no significant order effect for these problems; i.e., problems with a zero to the left of the multiplication sign were not responded to significantly faster or slower than problems with a zero to the right of the multiplication sign.

Ties.—For simple addition, Parkman and Groen (1971) found that ties did not systematically differ from one another in mean reaction time. However, in the present study, there was a significant tendency for the multiplication ties to increase in reaction time with an increase in the size of the multiplier (or multiplicand). An estimate of the size of this increase in reaction time with each additional increase in multiplier size, based on linear regression and pooling over positive and negative instances, was 10.1 msec. per increment, $F\ (1, 8) = 15.10$, $p < .005$. The overall mean reaction time for the ties, again pooling over positive and negative instances, was 600.7 msec.

DISCUSSION

The procedure used to perform the task required in this study can be viewed as having four successive stages: a stimulus preprocessing stage, a stage in which the product is computed or retrieved (multiplication stage), a stage in which the computed or retrieved product is compared to the third number (comparison stage), and a response execution stage. Of these four stages, only the second and third will be discussed in this section as the effects of the stimulus preprocessing and response execution stages were assumed to be constant throughout the study, except for practice effects, and to be noninteractive with the multiplication and comparison stages.

The primary purpose of this study was to investigate the processes underlying simple mental multiplication. However, the task used also involved a comparison operation. An important question then becomes to what extent can the effects of this comparison operation be separated from multiplication effects?

To answer this question, it is important to note that throughout the study there was a consistent finding that no matter how the reaction times were partitioned and analyzed, negative-instance means were always longer (usually by approximately 50 msec.) than positive-instance means. However, as can be seen from the results of the regression analyses performed for Table 3 and Table 4, the slope estimates from linear regression for the various structural variables investigated were very similar both for positive- and negative-instance data. In other words, the data for positive and negative instances have a consistent intercept difference, but similar slopes, when viewed as a function of the various structural variables investigated.

The finding that the regression lines for positive and negative responses differ in intercepts but not in slopes is evidence that the variables determining the slopes are affecting a different stage than are the variables determining the difference between intercepts, and that the processing durations of these respective stages are related in an additive manner (Sternberg, 1969). The variables determining these slopes (e.g., Sum and Min) are structural variables in multiplication. Hence, there is a compelling reason for associating these slopes with the multiplication stage mentioned above. Similarly, the variable determining the intercept differences is the relation (same–different) between $p \times q$ and r. It seems clear that this variable is associated with the comparison stage mentioned above, although an association with the response execution stage cannot be completely ruled out.

The comparison stage of the task used in this study can be viewed as a same–different matching operation. The finding that "same matches" require less time than "different matches" is consistent with previous research (e.g., Nickerson, 1965), which has shown that same matches for simple material are often 50–100 msec. faster than different matches.

With regard to the multiplication stage, the most interesting finding is the high degree of similarity between the pattern of reaction times found by Parkman and Groen (1971) for simple addition and that found for simple multiplication in the present study. In particular, both of these studies found a strong relationship between reaction time and the Min or Sum. This finding is consistent with the hypothesis that the performances of simple addition and simple multiplication are based on common underlying processes. If one assumes that simple multiplication is performed by a memory retrieval operation, then this close connection between multiplication and addition seriously questions the counting-model explanation of simple addition offered by Parkman and Groen. That is, simple addition would now also appear, at least in adults, to be based upon a memory retrieval operation.

Given a memory retrieval explanation of simple addition and multiplication, how are the reaction time effects associated with both operations to be interpreted? A number of possibilities exist which, due to the fact that very little data are available as to just how information is accessed and retrieved from long-term memory, must at this time be offered as conjectures only.

One possibility, briefly mentioned earlier, is that these effects are associated with the calculation of indexes used for memory retrieval. Such indexes, or cues, might either provide the memory location (address) of a desired number fact or might be used in a content-addressable fashion to locate the desired fact directly. The counting model of addition proposed by Parkman and Groen (1971) can easily be modified to encompass a memory retrieval operation and still be consistent with their data if it is assumed that (*a*) the model generates a retrieval index, rather than a sum, and that (*b*) the time for a memory retrieval operation, using the index so generated, is independent of the actual index used. The reaction time differences among the various simple additions would thus stem only from the differences in the times to set up retrieval indexes for these problems. In the same way, solutions to simple multiplication problems may be retrieved by use of such an index, calculated in the same manner as for simple addition, i.e., calculated in the manner described by the Parkman and Groen counting model. Again the time needed for a memory retrieval operation, after the index has been calculated, is assumed to be independent of the actual index used. Thus simple addition and simple multiplication, both based on common indexing and retrieval mechanisms, would be expected to display very similar reaction time effects, a result indicated by the present study and the earlier study of simple addition.

Another possibility is that simple addition and simple multiplication reaction time effects result from the memory retrieval operation itself. For example, perhaps the addition and multiplication "tables" are stored as hierarchical networks in long-term memory and the time to retrieve a given solution is a function of that solution's location in the network (due, perhaps, to linear, terminating searches through the nets). If this is true and if the networks for addition and multiplication are similar, then similar patterns of reaction times for performing the two types of operations would be expected.

The long latencies of problems in which either p or q equaled zero may be attributable to *S*s handling these problems in a manner quite different from that used for the other problems. In solving Min 0 problems, *S*s may have (unconsciously) used a rule similar to: "If a zero appears in the problem, generate a zero for the answer," while using a different process (perhaps memory retrieval, as suggested above) for non-Min 0 problems. Such an "explanation" is, at present, only conjecjecture but is at least consistent with the fact that teachers often introduce the "zero multiplication facts" later than the other basic problems of multiplication.

Apparently, reaction times for multiplication ties do not systematically differ from non-tie problems, at least not in the manner found by Parkman and Groen (1971) for addition. Hence, in multiplication it may be that ties and most non-ties are performed by means of identical processes, whereas in addition such problems may well be performed quite differently from one another.

Further evidence of the close relation between simple addition and multiplication reaction times has been recently reported by Parkman and Cavanagh (1971). Using a within-*S*s design, they found that the mean reaction time difference between simple addition and multiplication amounted to only 19 msec., with multiplication requiring the longer latency. In fact, for the six *S*s tested, no *S* deviated from this mean difference by more than 18 msec. Hence it would appear that simple addition and multiplication are closely related not only in pattern of mean reaction times, but also in performance duration.

REFERENCES

Dansereau, D. F. An information processing model of mental multiplication. *Dissertation Abstracts International*, 1969, **30** (4-B), 1916.

Dansereau, D. F., & Gregg, L. W. An information processing analysis of mental multiplication. *Psychonomic Science*, 1966, **6**, 71–72.

Groen, G. J., & Parkman, J. M. A chronometric analysis of simple addition. *Psychological Review*, 1972, **79**, 329–343.

Nickerson, R. S. Response times for "same-different" judgments. *Perceptual and Motor Skills*, 1965, **20**, 15–18.

Parkman, J. M., & Cavanagh, J. P. A model for simple numerical processing. Paper presented at the Mathematical Psychology Meetings, Princeton, N. J., September 1971.

Parkman, J. M., & Groen, G. J. Temporal aspects of simple addition and comparison. *Journal of Experimental Psychology*, 1971, **89**, 335–342.

Sternberg, S. The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 1969, **30**, 276–315.

Suppes, P., & Groen, G. J. Some counting models for first grade performance data on simple addition facts. In J. M. Scandura (Ed.), *Research in mathematics education*. Washington, D.C.: National Council of Teachers of Mathematics, 1967.

Thomas, H. B. G. Communication theory and the constellation hypothesis of calculation. *Quarterly Journal of Experimental Psychology*, 1963, **15**, 173–191.

(Received March 3, 1972)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 445–448

SHORT REPORTS

TIME ESTIMATION AND THE INTERSTIMULUS INTERVAL FUNCTION IN CLASSICAL CONDITIONING[1]

KURT J. TELLER, RICHARD DIETER, AND MILTON D. SUBOSKI[2]

Virginia Polytechnic Institute and State University

The *S*s estimated time by pressing a key just prior to the onset of the second of two lights. The interval between the lights was .2, .4, .8, 1.6, 3.2, 6.4, or 12.8 sec. for groups of *S*s. Accuracy of time estimation was best at the .4-sec. interval. In a second experiment, time estimation was good for a .6-sec. interval, poor for a 2.5-sec. interval. Time estimation remained good for a group which began estimating at the .6-sec. interval and was then progressively shifted to the 2.5-sec. interval. The results were interpreted as supporting the hypothesis that time estimation processes are important in determining interstimulus interval effects in classical conditioning.

Prokasy (Prokasy, 1965; Prokasy, Ebel, & Thompson, 1963) has suggested that time estimation processes are important in the acquisition of classical conditioned responses (CRs). The time estimation hypothesis is based on the assumption that *S* must time the CR to occur close to the UCS so that the reinforcing properties of CR–UCS overlap are maximized. According to the hypothesis, the decreasing levels of conditioning that are obtained when the interstimulus interval is increased beyond the optimal value are the product of the increasing difficulty of time estimation. This hypothesis receives support from investigations of time estimation. Woodrow (1951) reported that time estimation is most accurate for durations of approximately .6 sec., with increasingly less accurate estimation for longer or shorter durations. Many investigations of classical conditioning have reported optimal conditioning at CS–UCS intervals close to .6 sec. (e.g., Kimble, 1961). Another factor that could account for at least part of the poorer conditioning at interstimulus intervals below the optimum is simple reaction time (RT). Mean RTs are on the order of 140–180 msec. and distributions of RTs extend beyond 200 msec. (Woodworth & Schlosberg, 1954). Thus, interstimulus intervals below 200 msec. would be increasingly detrimental to the occurrence of CR–UCS overlap.

Experiment I was intended to investigate directly the time estimation hypothesis using a procedure analogous to the classical conditioning paradigm. Investigations of time estimation have typically involved judgements of the relative lengths of two intervals or the reproduction of an interval (Woodrow, 1951). In Exp. I, *S*s were instructed to press a key just prior to the end of an interval marked by the successive onset of two lights.

Prokasy et al. (1963) gradually shifted the interstimulus interval from a value (.6 sec.) yielding near-optimal levels of eyelid conditioning to a distinctly nonoptimal value (2.5 sec.). They found that the percentage of conditioned responses by this group remained equivalent to a group conditioned at the .6-sec. interval throughout. Furthermore, when the .6-sec. group was abruptly switched to the 2.5-sec. interval at the end of training, performance dropped immediately to a level equivalent to that of a group conditioned at the 2.5-sec. interval throughout. These results were interpreted (Prokasy, 1965; Prokasy et al., 1963) in terms of "shaping" appropriate time estimation in the shifted interstimulus interval group and a failure of accurate time estimation to develop in the abruptly shifted group. The results of Prokasy et al. (1963) provide for a powerful test of the time estimation hypothesis using the time estimation analogue to classical conditioning. Experiment II was intended to determine whether time estimation could be "shaped" in the same way as were the classical CRs of Prokasy et al.

Method.—The *S*s were 50 male and 50 female students in introductory psychology at Virginia Polytechnic Institute and State University, participating in the study for extra class credit.

The *S* was seated at a table with a flat-black plywood panel in front of him. The panel was 77 cm. wide, 31 cm. high, and contained two neon bulbs (NE 51), 77 mm. apart, centered 30 cm. above the base. A telegraph key was mounted 22 cm. directly below the lights. The light durations were controlled by a Lafayette timer (Model 504A), and response latencies were recorded on a Lafayette stop clock (Model 50807). The *E* was seated behind the panel and manually initiated trials with the toggle switch on the timer, which made a soft click, audible to *S*.

In Exp. I, five male and five female *S*s were randomly assigned to each of seven different interstimulus groups. The interstimulus intervals were .2, .4, .8, 1.6, 3.2, 6.4, and 12.8 sec. The interstimulus interval was the interval between the onset of the left light and the onset of the right light. The right light was always on during the intertrial interval. At the start of a trial, the right light went out and the left light simultaneously came on. The left light was on for the duration of the interstimulus interval and then went out and the right light simultaneously came on. Each *S* received 100 trials at intertrial intervals selected randomly from 10, 15, and 20 sec.

---

[1] This research was supported by funds from the Department of Psychology of Virginia Polytechnic Institute and State University and by Grant A0244 from the National Research Council of Canada.

[2] Requests for reprints should be sent to Milton D. Suboski, who is now at the Department of Psychology, Queen's University, Kingston, Ontario, Canada.

At the start of the session, each *S* was read the following instructions:

> Please listen carefully to the following instructions. You are participating in an experiment on time estimation. The panel in front of you contains two small lights and a key. Most of the time the right-hand light will be on. From time to time the right-hand light will go out and the left-hand light will come on. Your task is to watch the two lights carefully and press the key firmly just before the right-hand light comes on each time. It is important that each time you press the key, you hold the key down until the right-hand light comes on. Remember, each time the left-hand light comes on, you are to attempt to press the key just before the right-hand light comes on. You will have made a correct response each time that you have pressed the key as close as possible to the time when the right-hand light comes on. You will have made an incorrect response if the right-hand light comes on before you have pressed the key. Do you understand what you are to do? Okay, please tell me what you are supposed to do. Remember, it is very important that you hold the key down each time that you press it until the right light has come on. Okay, here we go.

In Exp. II, five male and five female *S*s were randomly assigned to each of three groups. The general procedure and instructions of Exp. I were used. Group S received 180 trials with interstimulus intervals as follows: 20 trials at .6 sec., 20 at .8, 25 at 1.0, 25 at 1.25, 30 at 1.6, 30 at 2.0, and the last 30 trials at a 2.5-sec. interval. Group 2.5 received 180 trials at a 2.5-sec. interval, and Group .6 had a .6-sec. interval for the first 150 trials and the last 30 trials at a 2.5-sec. interval. The interstimulus interval values and numbers of trials correspond to those used by Prokasy et al. (1963). At each interstimulus interval change, *S*s in Group S were given a break of about 1 min. while *E* changed timer settings. Similar breaks were given at the same points in the session to *S*s in the other two groups.

Results of Exp. I.—Figure 1 displays percentage of "correct" responses as a function of interstimulus interval for two different criteria for defining a correct response. The upper curve includes all responses within the interstimulus interval and shows that except at the .2-sec. interval, responses occurred within the interstimulus interval on approximately 90% of the trials at all intervals. The lower percentage of responses within the interstimulus interval at the .2-sec. interval probably reflects the truncation of the distribution of reaction times at this interval. The lower curve shows the function when only responses within .4 sec. of the end of the interval were scored as correct. This function shows that the accuracy of time estimation sharply decreased beyond the .4-sec. interval. Analysis of variance was conducted on percentage correct responses in blocks of 20 trials for the .4-sec. criterion. This analysis revealed significant effects of interstimulus interval, F (6, 63) = 14.32, trial blocks, F (4, 252) = 16.78, and the interaction, F (24, 252) = 2.03, all $p\text{s} < .01$. The effect of trial blocks reflected the general improvement in accuracy across trials. The interaction resulted from the fact that improvement in accuracy was greatest at the intermediate (1.6-, 3.2-, and 6.4-sec.) intervals.

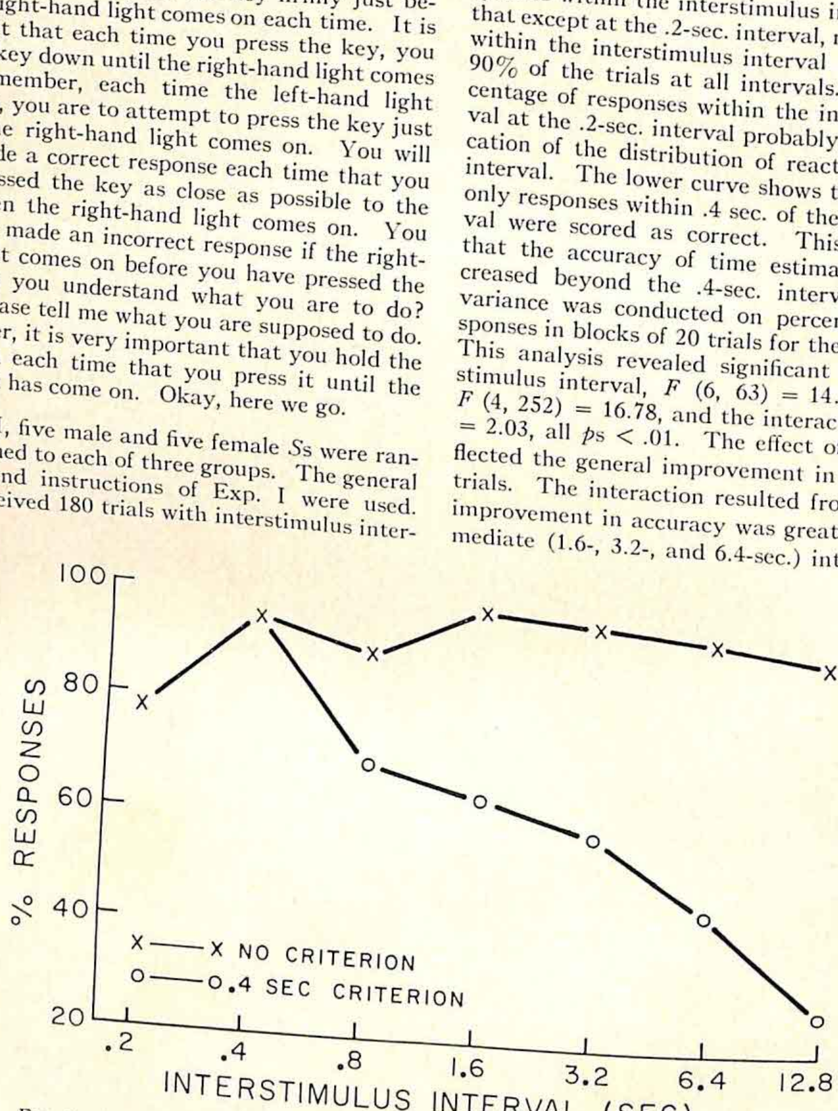


FIG. 1. Percentage of responses falling within the interstimulus interval (no criterion) and within .4 sec. of the end of the interstimulus interval (.4-sec. criterion) as a function of interstimulus interval.

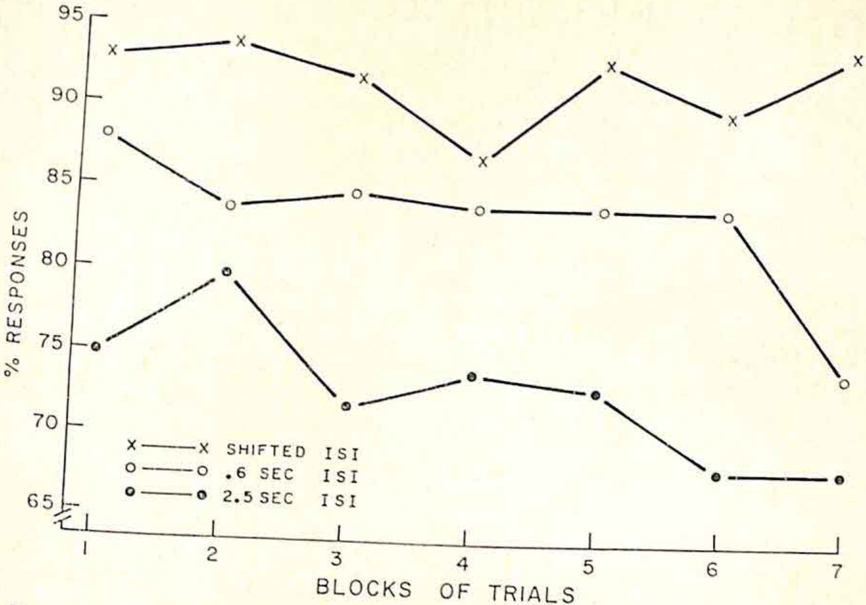


FIG. 2. Percentage of responses between .17 sec. after the start of the interstimulus interval and the end of the interstimulus interval. (The interstimulus interval was increased between each block of trials for Group S (shifted ISI). The interstimulus interval was increased for Group .6 (.6-sec. ISI) only between Trial Blocks 6 and 7. Group 2.5 (2.5-sec. ISI) received only a 2.5-sec. interstimulus interval.)

The mean standard deviation of response latency for each interstimulus interval group as a function of trial blocks was also examined (not shown). The variability of response latencies was an increasing function of interstimulus interval, $F\ (6, 63) = 18.35$, and decreased across trial blocks, $F\ (4, 252) = 14.03$, with the amount of decrease an increasing function of interstimulus interval, $F\ (24, 252) = 2.36$, all $p\text{s} < .01$.

Results of Exp. II.—To correspond to the Prokasy et al. (1963) procedure, correct responses were defined as those occurring between .17 sec. after the beginning of the interstimulus interval and the end of the interval. Figure 2 presents the percentage of correct responses for the three groups as a function of trial blocks. The trial block groupings of Fig. 2 are the same as those at which interstimulus interval shifts were made for Group S. Examination of Fig. 2 shows that Group S performed at a slightly higher level than Group .6 during the first trial block. This is a chance difference since the treatment of the two groups was identical during the first trial block. However, in spite of the progressive increase in interstimulus interval received by Group S, their accuracy of time estimation remained superior to that of Group .6, and both groups were considerably superior to Group 2.5. Analysis of variance of the data from the first six trial blocks revealed only a significant groups effect, $F\ (2, 27) = 3.92$, $p < .05$. Both trial blocks and the interaction yielded F ratios near unity. Analysis of the groups effect by orthogonal comparisons indicated that Group S and Group .6 did not differ ($F = 1.04$), but that Group S and Group .6 combined were significantly different from Group 2.5, $F\ (1, 27) = 6.81$, $p < .05$.

The accuracy of time estimation by Group .6 dropped substantially after the shift to the 2.5-sec. interval. Analysis of variance conducted on just the seventh trial block data failed to yield a significant groups effect, $F\ (2, 27) = 3.22$. However, orthogonal comparisons showed that Group 2.5 and Group .6 combined differed from Group S, $F\ (2, 27) = 6.15$, $p < .05$, whereas Group .6 did not differ from Group 2.5 ($F = .27$).

Discussion.—Experiment I clearly showed that the accuracy with which *S*s can anticipatorily respond to a signaled event is a function of interstimulus interval, with the greatest accuracy between interstimulus intervals of .2 and .8 sec. These results are in agreement with previous investigations of time estimation (e.g., Woodrow, 1951) and supportive of Prokasy's (1965) suggestion that time estimation processes are important in determining the interstimulus interval function in classical conditioning of skeletally mediated responses. As required by the time estimation hypothesis, accuracy of time estimation was best at the interstimulus interval that yields the highest level of classical conditioning. Furthermore, the present finding that the standard deviation of time estimates was an increasing function of interstimulus interval is similar to the results of Ebel and Prokasy (1963) for the latencies of eyelid CRs.

In Exp. II, the effects on time estimation of gradually increasing the interstimulus interval were highly similar to those of Prokasy et al. (1963) for the eyelid CR of humans and those of Prokasy and Papsdorf (1965) for the nictitating membrane CR of rabbits. Clearly, the ability of *S*s to estimate time accurately at long interstimulus intervals can be "shaped" by beginning at an interval favorable to time estimation and gradually increasing the interval. The importance of the shaping procedure is shown by the fact that an abrupt increase in the interstimulus interval led to an immediate and substantial decrement in the accuracy of time estimation. This result is also similar to the effect of an abrupt increase in the interstimulus interval in human eyelid conditioning (Prokasy et al., 1963). The rather remarkable similarity between eyelid conditioning and time estimation effects strongly suggests that time estimation processes are an important determinant of classical conditioning levels, and particularly the interstimulus interval function. The fact that time estimation appears to be important in classical conditioning only very indirectly supports the hypothesis that the reinforcing properties of CR–UCS overlap are of primary importance in the acquisition of classical CRs. Nevertheless, it does not appear likely that any alternative theoretical formulation would have predicted the outcomes of the present experiments as clearly.

REFERENCES

EBEL, H. C., & PROKASY, W. F. Classical eyelid conditioning as a function of sustained and shifted interstimulus intervals. *Journal of Experimental Psychology*, 1963, **65**, 52–58.

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning*. New York: Appleton-Century-Crofts, 1961.

PROKASY, W. F. Classical eyelid conditioning: Experimenter operations, task demands, and response shaping. In W. F. Prokasy (Ed.), *Classical conditioning*. New York: Appleton-Century-Crofts, 1965.

PROKASY, W. F., EBEL, H. C., & THOMPSON, D. D. Response shaping at long interstimulus intervals in classical eyelid conditioning. *Journal of Experimental Psychology*, 1963, **66**, 138–141.

PROKASY, W. F., & PAPSDORF, J. D. Effects of increasing the interstimulus interval during classical conditioning of the albino rabbit. *Journal of Comparative and Physiological Psychology*, 1965, **60**, 249–252.

WOODROW, H. Time perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, Rinehart, & Winston, 1954.

(Received June 18, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 448–450

VICARIOUS REINFORCEMENT AND MODEL'S BEHAVIOR IN VERBAL LEARNING AND IMITATION[1]

KAREN JOYCE KAPLAN[2]

University of Toronto

The effects of vicarious reinforcement (VR) and model's behavior on *S*s' imitation of the model's performance and learning to give critical responses were investigated in a verbal learning situation similar to that used by F. H. Kanfer and A. R. Marston in 1963. In acquisition, *S* said single words in alternation with a tape-recorded model. The model was either reinforced or was not reinforced for saying critical words, and he either increased, decreased, or held constant his critical response rate. No model was heard during an operant period which preceded acquisition or in extinction following acquisition. Results indicate that *S*s have a strong tendency to imitate a model, whether or not the model is reinforced or is successfully learning the task. The model's list content rather than VR accounts for *S*s' change in critical response rate during acquisition.

In a verbal conditioning study, Kanfer and Marston (1963) investigated the effects of vicarious and direct reinforcement (VR and DR) on *S*s' frequency of emitting critical responses (human nouns). The *S*s gave a word on signal in alternation with the tape-recorded voices of models (Ms) who were, in some conditions, reinforced (VR) with *E*'s spoken "Good" following critical responses (CRs). DR consisted of the word "Good" spoken by *E* when *S* said a critical word. The *S*s in four experimental groups heard tape recordings in which the frequency of critical responses increased with VR, DR, both VR and DR, or neither one (NR). The *S*s in two control groups heard tapes in which the frequency of CRs remained low and constant throughout, one with DR and one with NR, while two more control groups heard no tape recording but simply gave words on signal, again one with DR and the other with NR.

It was found that VR greatly facilitated learning in this task: Percentage of critical responses of experimental VR groups was at least twice that of the other experimental groups, both during the period of alternation with the recorded Ms and during an extinction period which followed. The addition of DR did not significantly enhance this effect. The group which heard the same tape recording with NR also showed an increased CR rate, showing that there is some effect due simply to M's CR rate; this increase, however, was much smaller than were those of the VR groups.

In a subsequent study, Marston and Kanfer (1963) demonstrated that while the number of Ms had no effect, the proportion of VR was an important determinant of *S*s' CR frequency. The omission of NR conditions, however, made it impossible to verify their earlier finding that *S*s' increased CR rate was due to VR rather than simply to M's CR rate. Such a verification was attempted by Phillips (1968), who replicated two of the conditions used by Marston and Kanfer (1963) with the addition of a control for imitation (defined as an increase in CRs attributable solely to the observation of a nonreinforced M whose CR rate increases). Phillips found no difference between VR and NR groups; *S*s imitated M regardless of whether or not M was reinforced.

---

[1] The author wishes to thank sincerely Ronald Cohen for his advice and assistance, given throughout the course of this research.

[2] Requests for reprints should be sent to Karen Joyce Kaplan, 75A Bristol Avenue, Toronto 170, Ontario, Canada.

One aim of the present study, then, is to gather further evidence either to support or to refute the original Kanfer and Marston (1963) finding that VR following M's CRs facilitates an increase in *S*s' CR rate.

In another study, Marston (1964) varied conditions in extinction following an acquisition period in which M's CRs increased and were each followed by VR. In extinction there were four conditions: In two groups, M's CR rate gradually decreased to its original level, one with VR continuing and the other with NR; in the other two, M's CR rate dropped immediately to a low level and remained constant, again with VR in one case and NR in the other. Thus for those *S*s who heard VR in extinction, the differences between acquisition and extinction lay solely in a change in M's CR rate. Marston found that in extinction all *S*s matched M's CR rate, that is, VR in extinction had no effect. Marston suggested that the preliminary acquisition period (with increasing M CR rate and VR) was necessary to establish an imitation set which, once established, is not weakened by withdrawal of VR.

The result obtained by Phillips (1968), however, puts such an intepretation in some doubt, since Phillips' control *S*s followed M in increasing their CR rate with neither concurrent nor previous VR. Thus the second aim of the present experiment is to expose *S*s to conditions similar to those in the extinction phase of Marston's (1964) study (i.e., with M's CR rate decreasing or constant, with or without VR) without the previous experience of his acquisition phase (M's CR rate increasing, with VR).

Method.—Tape recordings with an increasing, decreasing, or constant CR rate were used, with the prerecorded CRs either reinforced or not. Thus there were six conditons: three types of list (I, D, C) × two levels of reinforcement (VR, NR). Within each condition, half of the *S*s heard one set of words spoken by a male M, while the other half heard a second set of words spoken by a female M. For all VR conditions, the reinforcer was "Good" spoken by *E*'s voice following M's CR on the recording.

The words used in the recordings were taken from Thorndike and Lorge (those having a frequency of at least 50 per million). These were divided into two categories, human nouns and other words, discarding ambiguities, proper nouns, contractions, articles, and prepositions. Each word list consisted of 100 words, of which 55 were human nouns. The I lists consisted of words from the two categories, chosen randomly within the constraint that 1 human noun appear in the first block of 10 words, 2 in the second block, and so on, with the last block containing human nouns exclusively. The position of the human nouns within a block was randomly determined. The D lists were simply the *I* lists reversed. Each C list consisted of the words from the corresponding I list randomly scrambled, with the constraint that either 5 or 6 human nouns appear in every string of 10 consecutive words.

Instructions to *S* were recorded at the beginning of each tape, after which *S* was signaled to say a word once every 6 sec. for 200 trials. During the acquisition period (the middle 100 trials), M's intervening word was recorded 3 sec. before *S*'s signal, followed by VR when appropriate. The extinction period was followed by further instructions (to fill out a questionnaire) and the termination of the experiment.

The tapes were played to *S* on a Sony stereo tape recorder (Model TC-252) via earphones. His responses were tape recorded and also recorded in writing by *E*.

Ten *S*s were used in each condition. A total of 62 *S*s participated in the experiment, but protocols from 2 of them were discarded because they emitted fewer than 95% of the required responses. All *S*s were undergraduate psychology students at the University of Toronto.

Once settled in a small darkened room, *S* was instructed that he was participating in a "word study," that he was to say any single word that came to mind each time he heard the signal, which would be once every 6 sec., but that he should avoid saying phrases or proper names. He was also told that part of the time he would be "taking turns with someone else who will be going the same thing," and that during this time he would hear one of his (her) words through the earphones before each of his own signals.

At the end of the 200-response sequence, *S* was asked to turn on the light and to answer a questionnaire. To the five questions used by Marston and Kanfer (1963) were added three more concerning his perceptions of M's behavior and of any conflict he felt in choosing his own words. After the questionnaire was completed, *S* was told the nature and purpose of the experiment and any questions he had were answered.

Results and discussion.—Each *S*'s protocol was scored in 20 blocks of 10 response opportunities each. Failures to respond and ambiguous responses (i.e., responses which might or might not be intended as human nouns, e.g., son-sun, aunt-ant) were discarded, and each block was scored as a proportion of human nouns out of all remaining words. There was no systematic variation in the number of responses not given, and only 2% of the responses were missed or discarded. Table 1 shows the six group mean proportions for each block of 10 responses, except that the 50 operant trials are combined into one block.

Three separate analyses of variance were performed, using the arc sine transformation of each *S*'s CR proportions, to test for effects of list type,

TABLE 1

Mean Proportion of CRs by Experimental Condition

Group	Operant (50 trials)	Acquisition blocks										Extinction blocks				
		1	2	3	4	5	6	7	8	9	10	1	2	3	4	5
I, VR	.06	.11	.14	.16	.15	.23	.25	.34	.33	.20	.34	.12	.15	.17	.15	.09
I, NR	.10	.16	.13	.10	.13	.17	.23	.22	.32	.32	.43	.08	.13	.05	.06	.04
D, VR	.04	.32	.20	.39	.28	.20	.22	.13	.14	.10	.11	.13	.13	.07	.10	.06
D, NR	.06	.42	.25	.37	.27	.24	.17	.12	.06	.06	.11	.08	.03	.07	.05	.01
C, VR	.05	.20	.24	.28	.29	.19	.20	.26	.26	.26	.19	.09	.05	.15	.09	.12
C, NR	.06	.18	.29	.38	.32	.25	.32	.23	.30	.28	.24	.10	.09	.07	.07	.05

Note.—For acquisition blocks, 10 trials each; for extinction blocks, 10 trials each.

reinforcement, and blocks in acquisition and in extinction, and to test for effects of list type and reinforcement on changes in overall CR rate between operant and extinction periods. In each of the six conditions, approximately 23% of the unambiguous words emitted during acquisition were human nouns, with no significant differences among groups. The distribution of CRs over acquisition blocks in each condition resembled the distribution followed by M for that condition, although, since *S*s gave fewer CRs than did their Ms, the group CR rates fell far below Ms' levels. (Overall, *S*s gave 42% as many CRs as did Ms. Inspection of the graphs presented by Kanfer and Marston [1963] and by Marston and Kanfer [1963] suggest that similar measures for their *S*s all fall between 40% and 60%. Phillips [1968], on the other hand, had *S*s giving 165% the number of taped CRs in one experiment and 70% in another.) The single significant difference among distributions proved to be the difference during acquisition attributable to M's change in CR rate (List Type × Blocks interaction), F (18, 531) $= 6.71$, $p < .0005$. There were no significant differences in extinction; VR did not affect performance either in acquisition or in extinction.

This finding corresponds to that of Phillips (1968), who obtained an imitation effect but no VR effect, putting in further doubt the original Kanfer and Marston (1963) result. Like Phillips' result, it suggests that Marston (1964) was wrong in suggesting that an acquisition period with VR is necessary for the establishment of an imitation set. (Note that Marston's extinction period is to be compared with the present acquisition period: Both have D and C list types crossed with VR versus NR. The present study omits a preceding period of I list type with VR in order to determine whether such a period is necessary in order to establish imitation in the absence of VR.)

The questionnaires revealed that *S*s saw themselves as associating rather than as imitating; but these two strategies cannot be objectively distinguished, given the critical class used in this as in earlier experiments, because while common associations to human nouns are related human nouns (e.g., father–mother, painter–artist), other words elicit human noun assciations only rarely. As all experiments thus far have used human nouns, this simply suggests that we do not know whether the strong imitation effect reported here and by Phillips (1968) is independent of the critical category used.

REFERENCES

Kanfer, F. H., & Marston, A. R. Human reinforcement: Vicarious and direct. *Journal of Experimental Psychology*, 1963, **65**, 292–296.

Marston, A. R. Variables in extinction following acquisition with vicarious reinforcement. *Journal of Experimental Psychology*, 1964, **68**, 312–315.

Marston, A. R., & Kanfer, F. H. Group size and number of vicarious reinforcements in verbal learning. *Journal of Experimental Psychology*, 1963, **65**, 593–596.

Phillips, R. E. Vicarious reinforcement and imitation in a verbal learning situation. *Journal of Experimental Psychology*, 1968, **76**, 669–670.

(Received July 12, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 451–452

ESCAPE MAINTENANCE UNDER SERIAL AND SIMULTANEOUS COMPOUND PRESENTATIONS OF SEPARATELY ESTABLISHED CONDITIONED STIMULI

DONALD J. LEVIS[1] AND HARVEY S. LEVIN
University of Iowa

D. J. Levis reported in 1971 that human *S*s presented a serial CS procedure (S_1 followed by S_2) of 800 msec. or longer in a conditioning paradigm produced slower finger-withdrawal avoidance responses than *S*s presented a nonserial CS procedure. Using a slightly different procedure, this study provided a replication and a further test of the generalization decrement hypothesis offered in the Levis study. The components of a nonserial simultaneous compound CS (S_1S_2) and a serial compound CS condition (S_1/S_1S_2) were separately conditioned via an escape procedure before being combined into their respective conditions. It was predicted that for the serial CS group the above procedure would attenuate or reverse the response latency difference noted previously between the two CS conditions. The data support the above prediction.

In a recent study, Levis (1971) compared the effects of serial and nonserial CS presentation in the context of a delayed avoidance finger-withdrawal conditioning paradigm. The nonserial CS manipulation consisted of a constant light stimulus that was presented for the duration of the CS–UCS interval. The serial CS manipulation consisted of two differently colored light stimuli which were presented sequentially, with the onset of S_2 occurring at the midpoint of the CS–UCS interval. Electric shock served as the UCS. At the longer CS–UCS interval lengths (800 and 1,200 msec.), the nonserial CS *S*s produced faster response latencies than the serial CS *S*s, while at the shorter CS–UCS interval lengths (300 and 600 msec.), the latency difference between CS conditions was eliminated.

The above finding was interpreted from a conditioning framework. It was hypothesized that the fear associated to the first half of a given CS–UCS interval is dependent upon at least three factors: (*a*) the temporal proximity of this CS segment to the onset of the UCS; (*b*) the stimulus similarity of the first segment to the second segment which was paired directly with the UCS; and (*c*) the intensity of the UCS. In the case of the longer CS–UCS intervals, Factors 1 and 2 were believed to be responsible for hindering the generalization of fear from Segment 1 to Segment 2 for the serial CS conditions. In other words, with long CS intervals, proportionally less fear would be expected to generalize during acquisition from the last half of the CS interval, which was associated with shock, to the first, when these two portions of the CS interval comprise different stimuli, as with the serial manipulations, than when they involve the same stimulus, as was the case for the nonserial CS manipulation. In the case of the shorter CS intervals, Factor 1 was reduced resulting in an enhancement of fear generalization from S_2 to S_1.

Although the shortening of the CS–UCS interval effectively reduced the response latency difference between serial and nonserial procedures, the hypothesis outlined above suggests that similar results would be obtained as a consequence of any operation which facilitated the conditioning of fear to S_1 in the serial condition. As a further test of this hypothesis, the present experiment employed a converging operation designed to enhance conditioning of fear to S_1 in the serial procedure. The enhancement procedure consisted of conditioning *S*s to S_1 and S_2 separately before combining the two stimuli serially in an instrumental avoidance situation. This procedure should augment the amount of fear to the separate components when combined in a serial order and thus diminish the response latency difference obtained at long intervals between serial and nonserial procedures.

Method.—The study to be reported involved a separate replication in order to insure reliability of results. The *S*s in Replication 1 were 64 male volunteer undergraduate students obtained from the introductory psychology course research pool at the University of Iowa. The *S*s in Replication 2 were 64 male voluntary undergraduate students paid $1.25 for their ½ hr. of service.

Each *S* was seated in a dimly illuminated room in front of a table upon which was mounted a panel board. Attached to the board were one red and one blue 115–125-v., 25-w. lamps horizontally arranged. The *S*'s response unit consisted of a telegraph key attached to an aluminum plate. Electric shock was generated by a 60-Hz. 120-v. ac source stepped up by a high-voltage transformer to 750 v. The shock was administered through a copper disk on the surface of the response key and the aluminum plate on which *S* rested his hand. Programming equipment and *E* were located in an adjoining room. A more detailed description of the apparatus can be found in Levis (1971).

The procedure closely paralleled that used by Levis (1971) for the simultaneous compound (S_1S_2) and serial compound (S_1/S_1S_2) CS procedures. The illumination of a red and blue light on the panel board served as the CS. The shock source, which provided the UCS, was set to produce 3.65 ma.

---

[1] Requests for reprints should be sent to Donald J. Levis, Department of Psychology, University of Iowa, Iowa City, Iowa 52240.

(rms.) of current. The CS–UCS interval was 800 msec. The *S*s were instructed to respond by removing the index finger of their dominant hand from the telegraph key. They were first told their task was to respond after shock onset. They were then given two escape trials which was the mean number of escape trials made by *S*s in the Levis experiment; next, they were instructed that their task was to avoid shock. Unknown to them, shock was discontinued at this point. These extinction trials continued until *S*s either responded on three consecutive trials with a latency of 2,800 msec. or reached a total of 100 extinction trials, whichever occurred first. A response in the extinction phase of the experiment immediately terminated the CS.

Four groups of 16 *S*s were tested during each replication. The CS manipulation for Groups 1 and 2 involved the same manipulation for both the escape and extinction phases of the experiment. Group 1 ($S_1S_2 - S_1S_2$) *S*s were exposed in both phases to the simultaneous presentation of two CS lights. Group 2 ($S_1/S_1S_2 - S_1/S_1S_2$) *S*s were presented a two-component serial compound CS for both the escape and extinction trials. The first stimulus, S_1, came on with the start of the trial and remained on until a response was made. The onset of S_2 occurred at the midpoint of the CS–UCS interval (400 msec.) and remained on with S_1 until a response was made. The *S*s in Groups 3 and 4 were exposed to a different escape procedure. Group 3 ($S_1, S_2 - S_1S_2$) *S*s were presented one of the CSs for one escape trial and the other CS for the second escape trial. In extinction they received the simultaneous presentation of both CSs on each trial. Group 4 ($S_1, S_2 - S_1/S_1S_2$) *S*s were presented one of the CSs for one escape trial and the other CS for the second escape trial. In extinction, the two CSs were presented in a serial compound order in the manner of Group 2. Both the lamp color and order of CS presentation were appropriately counterbalanced within each group for both the escape and extinction phases of the experiment. The left socket on the panel board always held the lamp which served as S_1 for the serial condition. The order of running *S*s was determined randomly.

Results.—One *S* in Replication 1 who requested the experiment be terminated after one shock trial was discarded and replaced. All *S*s experienced two shocks resulting in the following mean shock durations, combining replications, for Groups 1, 2, 3, and 4 of .192, .198, .199, and .190, respectively. None of the comparisons between groups or the Group × Replication interaction approached significance. However, the replication factor was significant, $F(1, 120) = 128.4$, $p < .001$. The *S*s in Replication 2 responded faster on escape trials.

The extinction indexes analyzed were (*a*) number of extinction trials and (*b*) mean response latency. In Replication 1, 4 *S*s in Group 2 and 2 *S*s in Group 3 failed to make one response latency less than 800 msec. The immediate extinction of these *S*s presents a problem in analyzing Index *b* since their inclusion spuriously inflates *S*s' latency data. This was not a problem in the Levis (1971) study since all *S*s made responses below 800 msec. before extinguishing, and the three criterion extinction trials (latencies over 2,800 msec.) were eliminated from the analysis. To provide a comparable analysis, both indexes of the present study were evaluated for the fastest 12 *S*s in each group (defined by their mean response latency) for each replication. The mean numbers of trials to extinction for Groups 1, 2, 3, and 4, combining replications, were 70.6, 72.8, 72.5, and 82.3, respectively. None of the groups reliably differed from each other, nor was the replication factor or Groups × Replication interaction significant. The lack of reliable differences between groups is not surprising in light of the finding that 74% of the *S*s analyzed failed to extinguish prior to the 100th trial. Sixteen *S*s in Group 1 and in Group 2, 18 *S*s in Group 3, and 21 *S*s in Group 4 were terminated after Trial 100.

The mean extinction latencies for Groups 1, 2, 3, and 4, combining replications, were .355, .469, .262, and .289 msec., respectively. Separate planned comparisons between Groups 1 and 2, 1 and 3, and 2 and 4 were made. Group 1 *S*s produced faster response latencies than *S*s in Group 2, $F(1, 88) = 4.94$, $p < .05$. Group 1 and 3 *S*s did not reliably differ in response latencies, while Group 4 *S*s produced faster response latencies than Group 2 *S*s, $F(1, 88) = 12.38$, $p < .001$. The replication factor and Groups × Replication interaction did not approach significance. To insure the obtained between-group latency differences were not a function of those *S*s who extinguished quickly, an analysis was performed over the data from the fastest 16 *S*s in each group who had not extinguished by Trial 100. The means for Groups 1, 2, 3, and 4 were .286, .413, .234, and .225, respectively. Comparable results to the previous analysis were obtained. Group 1 reliably differed from Group 2 ($p < .001$), Group 4 reliably differed from Group 2 ($p < .001$), and the Group 1 versus Group 3 comparison was nonsignificant. Inspection of the data revealed that the group mean latency for these *S*s remained reasonably stable across blocks of 10 trials.

Discussion.—The comparisons between Groups 1 and 2 essentially provided a replication of the simultaneous and serial compound CS procedures tested in Levis (1971). The expected latency difference between the two groups was found, with the serial CS *S*s producing longer latency responses. The separate conditioning of S_1 and S_2 on escape trials for the serial condition (Group 4) did reliably attenuate the latency difference between the serial and nonserial CS procedures. However, the separate conditioning of S_1 and S_2 on escape trials for the simultaneous compound procedure (Group 3) did not reliably attenuate the latency differences for that procedure, although it is possible that Group 3 data may have been affected by a "ceiling" effect. The generalization interpretation made in the Levis study appears consistent with the above findings.

REFERENCE

Levis, D. J. Effects of serial CS presentation on a finger-withdrawal avoidance response to shock. *Journal of Experimental Psychology*, 1971, **87**, 71–77.

(Received August 7, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 453–455

THE INTERRELATIONSHIP BETWEEN THE EBBINGHAUS AND DELBOEUF ILLUSIONS

JOAN S. GIRGUS[1]
City College of the City University of New York

STANLEY COREN
New School for Social Research

MYRA AGDERN
City College of the City University of New York

Variation of the size of the context circles in the Ebbinghaus illusion results in a systematic variation of the apparent size of the central test figure that is in accord with a comparative judgmental model of this illusion. However, increasing the distance of the context figures results in a decrease in the apparent size of the test figure regardless of the size of the context circles. This result is not predicted by a judgmental model and indicates a similarity between the Ebbinghaus and Delboeuf illusions.

An explanation for the Ebbinghaus or Titchener circles illusion in terms of the operation of a comparative judgmental mechanism has recently been proposed by Massaro and Anderson (1971). According to their model, the size of the central circle is judged relative to the size of the nonconcentric surrounding or context circles. The contribution of the context figures is determined by their size and number, and weighted in accordance with their distance from the central circle. On the other hand, Cooper and Weintraub (1970) have suggested that the Ebbinghaus illusion (central circle surrounded by nonconcentric circles) is a variant of the Delbouef illusion (central circle surrounded by a concentric circle). They support their contention by demonstrating similar changes in illusion magnitude for these figures when there is a temporal delay between presentation of the central test circle and the surround. This second approach suggests that the ring of nonconcentric circles around the test circle in the Ebbinghaus illusion provides the perceptual configuration of a concentric circle. These two ways of looking at the Ebbinghaus illusion would make similar predictions in some, but not all, situations.

Let us consider the effect of varying the distance between the surrounding circles and the test circle. When Massaro and Anderson (1971) vary this parameter in the condition where the surrounding circles are smaller than the test circle, they find a decrease in the apparent size of the test circle as distance increases. This is in accord with their judgmental model. This finding is also in accord with a model based on the Delboeuf, where the usual finding is that the size of the inner circle decreases steadily as the size of the concentric outer circle increases (Oyama, 1960; Weintraub, Wilson, Greene, & Palmquist, 1969). However, the two approaches make opposing predictions in the Ebbinghaus situation in which the surrounding circles are larger than the test circle. In this situation, which was not tested by Massaro and Anderson, the judgmental model would predict an increase in the size of the test circle (a decrease in the illusion) as the distance between test and surrounding circles increases. On the other hand, prediction from the Delboeuf model would still indicate a decrease in the size of the inner test circle (an increase in the illusion) as the distance between the test circle and larger surrounding circles was increased.

A model of the Ebbinghaus illusion based on the Delboeuf would also predict both an assimilation and a contrast phase as a function of distance between the test circle and the "ring" of nonconcentric surrounding circles. As distance increases, the apparent size of the center circle should appear first larger than, then equal to, and finally smaller than a circle with no surround (Oyama, 1960). In the situation in which the two approaches both predict a decrease in test circle size (surrounding circles smaller than test circle), the judgmental model simply predicts that the apparent size of the center circle surrounded by smaller circles would decrease to equality, while the Delboeuf model predicts that the test circle will actually begin to show an illusion in the opposite direction as the distance continues to increase. In order to test this prediction, an estimate of the apparent size of the test circle without any surrounding circles is needed. This was not provided by Massaro and Anderson (1971).

The judgmental and the Delboeuf models of the Ebbinghaus or Titchener circles illusion make similar predictions when distance is varied in the situation in which the surrounding circles are smaller than the test circles but different predictions for larger surrounding circles. The experiment presented here varies both the size of the inducing circles and the distance between inducing and test circles in order to differentiate between the two proposed bases for the Ebbinghaus illusion. In addition, a test circle with no surrounding circles has been included in this experiment in order to check for the existence of both an assimilation and a contrast phase as the distance between test and surrounding circles is varied.

Method.—All figures were black outline circles. The central test circle was held constant at 14 mm. The test circle was either presented alone, or with four surrounding circles. The diameter of these context circles was 5, 10, 18, or 23 mm. The distance between the proximal edges of the central and context figures was 3, 6, 12, or 24 mm.

Responses were made by rotating a wheel that presented single comparison circles (with no sur-

---

[1] Requests for reprints should be sent to Joan S. Girgus, Department of Psychology, City College, City University of New York 10031.

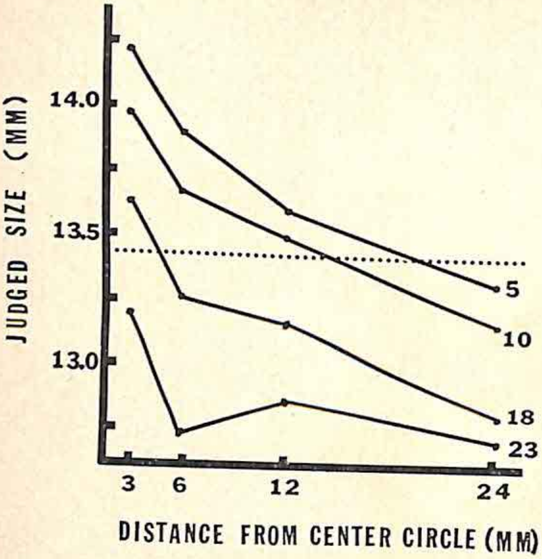


FIG. 1. Mean diameter of comparison circle judged equal to center circle of Ebbinghaus figure plotted as a function of distance from center to surrounding circles. (Size of surrounding circles in millimeters is listed at the right of each curve. The dotted line represents the judged size of a circle with no surrounding elements.)

rounding circles) ranging in diameter from 8.0 to 19.5 mm. in steps of .5 mm. Single circles appeared in a 26-mm. aperture cut into the apparatus.

Fifteen *S*s judged the apparent size of the central test circle once for each stimulus in the set, presented in randomized order. The *S*s were allowed free eye movements and could look back and forth between the stimulus and comparison figures.

Results and discussion.—Figure 1 presents the mean judged circle size for the 17 stimulus configurations used. The dotted line in the figure represents the mean judgment of the size of a circle seen in isolation. The departure from the expected 14 mm. is a constant error probably due to the presence of the aperture through which the comparison circles were viewed.

It is clear that, in general, smaller context circles result in the test circle appearing as larger, while larger context circles result in smaller size estimates for the test figure. The effect of circle size is significant, with $F\ (3, 42) = 22.10$, $p < .001$. The effect of the distance of the context circles upon the size estimate is also significant, with $F\ (3, 42) = 15.69$, $p < .001$.

The change in size estimates with increasing distance of the context circles helps shed some light on the underlying mechanisms for the Ebbinghaus illusion. For the smaller context circles it is clear that increasing the distance results in a decrease in the apparent size of the test circle, which replicates the findings of Massaro and Anderson (1971). However, for the larger context circles, we find exactly the same decrease in apparent size with increasing distance of the context circles rather than the opposite trend that would be predicted from a comparative judgmental theory. The similarity between the curves for both the larger and the smaller surrounds is supported by a lack of significant interaction between surrounding circle size and distance, $F\ (9, 126) = .47$. This decrease in apparent size of the test figure with increasing distance of the context circles, regardless of their size, is in accord with predictions which would be made on the basis of a similarity between the Ebbinghaus illusion and the Delboeuf, but is not in accord with predictions that would be made from a comparative judgmental model. A judgmental theory would predict a decrease in the size of the test circle with increasing distance of the context figures only for the smaller surrounding figures. The larger context circles should produce an increase in apparent size, as the increasing distance between test and inducing figures lowers the weighting of these figures in the comparative judgment.

Further evidence for the similarity of the Ebbinghaus and Delboeuf is illustrated by the fact that for the 10-mm. context circles at the farthest distance (24 mm.), the apparent size of the test circle is actually smaller than the same circle seen without any context figures, sign test, $p < .05$. This reversal from over- to underestimation with large distances between the test and context figures is consistent with the Delboeuf mechanism (Oyama, 1960) and difficult for a simple judgmental theory to handle.

Despite these apparent similarities between variations in the magnitude of the Ebbinghaus illusion and the Delboeuf, it is important to note that no significant amount of assimilation occurred at the closest distances for the 18-mm. and 23-mm. context circles and no significant amount of contrast occurred for the 5-mm. circles at the farthest distances. In addition, the variation in the size of the context circle results in clear apparent-size changes in the test figure, which is consistent with a judgmental theory. Such variation would be difficult to explain in terms of a Delboeuf-type mechanism unless one attempts to utilize the center of gravity of the surrounding context figures rather than the proximal distance as the measure of relative distance between center circle and surround.

Further evidence for a different judgmental component in the Ebbinghaus figure has recently been provided by Coren (1971), who has shown that the illusion can be demonstrated in configurations in which the context circles are physically equal in size to the test circle, but made apparently larger or smaller through the operation of size constancy in a stereoscopic paradigm. Thus, as has been suggested for many other classical visual geometric illusions (Coren, 1970; Over, 1968), the Ebbinghaus illusion appears to be multiply caused, and may even be a compound of several illusions rather than a unitary perceptual distortion.

REFERENCES

COOPER, L. A., & WEINTRAUB, D. J. Delboeuf-type circle illusions: Interactions among luminance, temporal characteristics, and inducing-figure variations. *Journal of Experimental Psychology*, 1970, **85**, 75–82.

COREN, S. Lateral inhibition and geometric illusions. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 274–278.

COREN, S. A size-contrast illusion without physical size difference. *American Journal of Psychology*, 1971, **84**, 565–566.

MASSARO, D. W., & ANDERSON, N. H. Judgmental model of the Ebbinghaus illusion. *Journal of Experimental Psychology*, 1971, **89**, 147–151.

OVER, R. Explanations of geometrical illusions. *Psychological Bulletin*, 1968, **70**, 545–562.

OYAMA, T. Japanese studies of the so-called geometrical-optical illusions. *Psychologia*, 1960, **3**, 7–20.

WEINTRAUB, D. J., WILSON, B. A., GREENE, R. D., & PALMQUIST, M. J. Displacement versus diameter, arc deletions, and brightness contrast. *Journal of Experimental Psychology*, 1969, **80**, 505–511.

(Received August 13, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 455–457

FREE RECALL FOLLOWING A SWITCH IN ENCODING CLASS[1]

MICHAEL S. HUMPHREYS[2]
University of British Columbia

WILLIAM M. PETRUSIC
Carleton University

ROBERT M. SCHWARTZ
University of British Columbia

Following a Brown-Peterson task in which taxonomic categories were shifted on Trial 5 for the experimental condition but were not shifted for the control condition, *S*s were asked to free recall as many words from all five trials as they could. Results of the free-recall test showed a reduction of the difference between experimental and control conditions for recall of Trial 5 items. There was no difference between experimental and control conditions in free recall of items presented on the first four trials.

In the Brown-Peterson paradigm (Brown, 1954; Peterson & Peterson, 1959), the presentation and/or learning of the first few items interferes with performance on the later items. This buildup of proactive interference (PI) occurs rapidly and reaches a maximum after three or four trials (Keppel & Underwood, 1962). Wickens, Born, and Allen (1963) have demonstrated that the interference may be specific to the class of materials used, such that switching the class of materials after a few trials leads to improved performance on the trial. Wickens (1970) has reviewed a number of studies which have demonstrated this "release from PI," and he assumes that a switch in materials which leads to a release from PI results from a switch in encoding class.

The present study was designed to provide information about what has been learned following a switch in encoding class by asking *S*s to free recall all of the items immediately after the test for retention of the last set of items. The free-recall task should provide evidence about: (*a*) the extent to which the release from PI is due to increased list differentiation as opposed to reduced interference between items (cf. Peterson, 1967) and (*b*) whether the release from PI serves only to improve performance on the "release" items or whether it affects recall of the earlier items. Data relevant to the first point are the free-recall performances of experimental and control *S*s on items that were presented on the release trial as compared to the immediate test performance on these items. Data relevant to the second point are the free-recall performance of experimental and control *S*s on items that were presented prior to the release trial.

A previous, unpublished experiment had indicated that free recall of items presented prior to the release trial was not better, and was possibly worse, following a shift in encoding class than following no shift. The present study was designed as a replication, with appropriate controls and better procedures, of that study.

Method.—Five triads of animal names and five triads of vegetable names were constructed from Bousfield's (1953) lists such that the mean frequency of occurrence between triads was approximately equal. Each triad was typed in a column on a 3 × 5 in. card.

Each animal triad was paired with a vegetable triad, and both triads within a pair were assigned a three-digit number which was used as a starting number for the counting-backwards task.

Order, materials, and conditions were varied between *S*s. The order variable was a Latin square arrangement of triads for presentation, such that each triad appeared equally often in each of five presentation positions. The Latin square arrangement was fitted to each set of materials, animals and vegetables, to form the control groups. For the conditions variable, experimental groups were defined by interchanging the fifth triad of animal and

---

[1] This research was supported by Grant APA 337 from the National Research Council of Canada. We thank Mary-Ann Starchuk for collecting the data.

[2] Requests for reprints should be sent to M. S. Humphreys, Department of Psychology, University of British Columbia, Vancouver 8, Canada.

TABLE 1

MEAN NUMBER OF WORDS RECALLED ON THE IMMEDIATE TEST AS A FUNCTION OF CONDITIONS AND TRIALS

Cond.	Trial				
	1	2	3	4	5
Experimental	2.33	2.13	1.48	1.45	2.18
Control	2.38	2.03	1.48	1.63	1.50

vegetable control groups which occupied the same row of the Latin square. Thus, half of the control *S*s were presented five animal triads and the other half were presented five vegetable triads, while half of the experimental *S*s were presented four animal triads followed by a vegetable triad and the other half were presented four vegetable triads followed by an animal triad. Combining order with materials and conditions resulted in 20 groups.

To assign *S*s to the 20 groups, the orders were randomly arranged and the combinations of conditions and materials within an order were randomly arranged. Thus, the first four *S*s served in the same order and one of the four served in each of the four combinations of conditions and materials within that order, etc. This entire procedure was repeated four times so that there were four *S*s in each of the 20 groups.

The *S*s were 80 male and female volunteers who responded to an advertisement in the student placement office; *S*s were tested individually and were paid for their participation. Following this experiment, *S*s also participated in another, unrelated experiment.

The *S* was first given practice in counting backwards by threes. The *E* read a three-digit number, saying each digit successively in time to a metronome which was beating at a 1-sec. rate. On the following beat, *S* started counting backwards by threes by responding with each digit successively also in time to the metronome. This procedure was repeated until *E* was satisfied that *S* could reliably perform the task.

For each trial of the experiment, *E* showed a triad to *S*. Starting with the first beat after the presentation of the traid, *E* read the three-digit number assigned to that triad. Upon reading the last digit, *E* removed the triad and *S* started to count backwards by threes. Thus, the triad was shown to *S* for 3 sec., but during the last 2 sec. *S* had to pay attention to the reading of the three-digit number. After 15 beats, *E* said "recall" and *S* attempted to recall the words in the triad. Approximately 6 to 8 sec. were allowed for recall before *E* initiated the next trial.

After the completion of the fifth trial, *E* handed *S* a sheet of paper containing 15 lines; *S* was asked to write down as many words from the experiment as he could remember. There was no time limit for this task.

Results.—A $2 \times 2 \times 4$ analysis of variance was used to examine performance on the first four triads. The factors were materials (animals vs. vegetables), conditions (experimental vs. control), and trials. As order was used only for counterbalancing, the order variable was not examined in this or subsequent analyses. The results of the analysis showed that the only significant effect was associated with trials, F (3, 228) = 22.81, $p < .001$. Performance of experimental and control conditions collapsed over materials are presented in Table 1. Examination of Table 1 shows that the significant trials effect indicates a decline in performance from Trial 1 to Trial 4.

A second analysis, a 2×2 analysis of variance, was used to examine Trial 5 performance. The only significant effect was that of conditions, F (1, 76) = 13.50, $p < .001$. As Table 1 shows, the switch in categories for the experimental *S*s resulted in improved Trial 5 performance.

Intrusion data were also examined. On Trial 2 through Trial 4, control and experimental groups made totals of 35 and 39 intrusions from previous trials, respectively. On Trial 5, the control group made 16 such intrusions while the experimental group made only 3 such intrusions. When expressed as a percentage of the number of Trial 5 errors made, intrusion errors constituted 9% of the errors in the experimental condition and 25% of the errors in the control condition.

The data from the free-recall task following Trial 5 are shown in Table 2. While there is still a large difference between experimental and control conditions in Trial 5 recall, the difference is smaller than in the immediate test for retention. The principal reason for this reduction is that control *S*s recalled 14 Trial 5 words which they had not recalled during the immediate test, while the experimental *S*s recalled only 2 such words. A chi-square analysis of the number of *S*s recalling at least 1 Trial 5 word which had not been recalled on the immediate test found this difference to be significant, $\chi^2(1) = 9.92$, $p < .005$.

In terms of the total number of words recalled from the first four triads, there was no effect of conditions. Control and experimental *S*s recalled totals of 212 and 213 words from the first four triads, respectively. Also there was no effect of materials, nor was there evidence for a Materials × Conditions interaction, all Fs < 1. The only F value which exceeded 1 was a nonsignificant effect of trials, F (3, 228) = 1.83, $p > .10$. Further analysis of the trials effect was made for all five trials of the

TABLE 2

FINAL RECALL EXPRESSED AS THE MEAN NUMBER OF WORDS RECALLED AND THE PERCENTAGE OF IMMEDIATE RECALL AS A FUNCTION OF CONDITIONS AND TRIALS

Measure	Cond.	Trial				
		1	2	3	4	5
Mean no. recalled	Experimental	1.05	1.38	1.45	1.45	2.08
	Control	1.20	1.38	1.30	1.43	1.58
% of immediate recall	Experimental	.45	.65	.98	1.00	.95
	Control	.51	.69	.88	.88	1.05

control group, but no significant effect was found, $F < 1$. The failure to find a significant effect of trials suggests that there was no recency effect in terms of the number recalled. Of course, if recall is expressed as a percentage of the amount recalled on the immediate test, there is a pronounced recency effect. Final recall expressed as the mean number of words recalled and the percentage of immediate recall is shown in Table 2.

Discussion.—The free recall of Trial 5 items showed a reduction of the effect of shifting encoding class as compared to the effect found in the immediate test. The percent recovery (the difference between the control and experimental conditions on Trial 5 divided by the difference between Trial 1 and Trial 5 control performances) for the immediate test was 77%, nearly the same as that found by Loess (1967) in a study which also shifted taxonomic categories. However, when the free recall data are used in place of the immediate test data, the percent recovery is 63%. This difference indicates that list differentiation could account for an apparent recovery of as much as 14%; the reduction of intrusion errors on Trial 5 of the immediate test for experimental *S*s and the improvement of free recall of Trial 5 items for control *S*s support this conclusion. Although the percent recovery that could be due to list differentiation accounts for only a small proportion of the recovery in the present data, list differentiation might be the most important component of the observed recovery of 20%–25% following shift in the number of syllables, number of phonemes, and articulatory dimensions (Wickens, 1970).

The free recall data also indicated that there was no difference between experimental and control groups in recalling items from the first four trials. It appears that shifting encoding classes affected performance on the shift items without affecting performance on the previous items. Thus, there appears to be some asymmetry of proactive and retroactive effects.

The failure to find a recency effect in the free recall data, except when expressed as a percentage of the amount recalled in the immediate test, suggests two possibilities. First, the immediate test might have provided more opportunity for rehearsal of earlier items since more early items are recalled in the immediate test. Alternatively, the buildup of PI may have hindered the learning, not just the immediate recall performance, of the later items.

REFERENCES

Bousfield, W. A. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 1953, **49**, 229–240.

Brown, J. The nature of set to learn and of intra-material interferences in immediate memory. *Quarterly Journal of Experimental Psychology*, 1954, **6**, 141–148.

Keppel, G., & Underwood, B. J. Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 153–161.

Loess, H. Short-term memory, word class, and sequence of items. *Journal of Experimental Psychology*, 1967, **74**, 556–561.

Peterson, L. R. Search and judgment in memory. In B. Kleinmuntz (Ed.), *Concepts and the structure of memory*. New York: Wiley, 1967.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **54**, 157–173.

Wickens, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

Wickens, D. D., Born, D. G., & Allen, C. K. Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 440–445.

(Received September 15, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 458–459

"STROOP" EFFECT: INPUT OR OUTPUT PHENOMENON?[1]

DOUGLAS L. HINTZMAN,[2] FRANK A. CARRE, VERONICA L. ESKRIDGE,
ANTHONY M. OWENS, STEPHANIE S. SHAFF, AND M. ELAINE SPARKS

University of Oregon

Two explanations of the Stroop color-naming effect are: (*a*) that the form of the word and the color of the ink contact memory in parallel, causing covert response competition when they elicit different color names and (*b*) that color words distract from the ability to input or to encode ink color. In a mixed-list design, latencies of color-naming responses were taken for individual stimuli falling into four categories: (*a*) In S, the word and color name were the same; (*b*) in D, the word named a different color; (*c*) in N, the word was neutral with respect to color; (*d*) in C, the stimulus was a nonword. Compared to Cond. C, S responses showed facilitation, D showed interference, and N showed no effect. The facilitation in Cond. S is as predicted by the hypothesis that the Stroop effect is due to response competition rather than encoding interference.

If *S* is instructed to ignore the meaning of a word and to name as quickly as possible the color of ink in which the the word is printed, the amount of time taken to name the color depends on the word's meaning. If the word names a color other than the color of ink, the latency is greater than if the word is neutral with respect to color.

In two recent papers, this phenomenon—the "Stroop" effect—has been given different interpretations. Keele (1972) has argued that selective attention must operate after, rather than before, memory retrieval. That is, the form and color of the stimulus are processed in parallel, and when these two sources of information elicit different color names, covert response competition results. This explanation localizes the Stroop effect at the output of the memory system. Hock and Egeth (1970), in contrast, localized the effect at input. Their explanation was that the nature of the task makes a color name particularly "attention-catching," so that it interferes with the ability to process and encode the color of ink.

A way to discriminate between these two interpretations is to include a condition in which the color word and the color of ink match (e.g., BLUE in blue ink). If the Stroop effect is caused by response competition, then this condition should not produce interference. In fact, since both the form and the color of the stimulus lead to the same response, the color should be named faster than that of a neutral control word. On the other hand, if color words interfere with the ability to encode ink color, then the proposed condition should produce interference like that obtained when the color word does not match the color of ink.

Traditionally, Stroop experiments have been done by manipulating conditions between lists (e.g., Pritchatt, 1968), a design that is subject to confounding by the use of different strategies in different conditions. The present experiment followed Keele (1972) in using a discrete-trial procedure and mixed-list design to avoid this problem. Unlike Keele's study, a vocal response was used. There were four conditions. In two of them the stimuli were color words that were either the same as or different from the required color-naming response. In a third the stimuli were neutral words, and in the control condition they were nonwords.

Method.—Stimuli were produced by printing each word on a white card in one of four colors of ink: RED, GREEN, BLUE, or GOLD (yellow-orange). The letters were $\frac{1}{4}$ in. square, and the lines approximately $\frac{1}{32}$ in. wide. The stimuli fell into four categories: Cond. Same (S), in which the word named the color of ink in which it was printed; Cond. Different (D), in which the word named one of the three other colors; Cond. Neutral (N), which used noncolor words matched with color names in terms of first letter and word length, i.e., RAT, GLASS, BIRD, GLAD; and Cond. Control (C), which used anagrams of the color names with the first letters unchanged, i.e., RDE, GENER, BEUL, GDLO. Stimuli of Cond. S and D were generated by pairing each color name equally often with each ink color. Thus, Cond. D occurred three times as often as S, and color name was not correlated with color of ink. Altogether there were 128 experimental cards: 16 in Cond. S, 48 in D, 32 in N, and 32 in C. The stimulus sequence was divided into four blocks of 32 stimuli each (the order of presentation of these blocks was varied systematically across *Ss*). Each block, in turn, was divided into four subblocks. Each subblock consisted of 1 card in Cond. S, 3 in D, 2 in N, and 2 in C, presented in random order.

Nineteen students in an undergraduate learning course at the University of Oregon served as *Ss* for extra course credit. The data of one *S* who made an unusually large number of errors and later confessed to being color-blind were discarded.

Each *S* was seated before a tachistoscope and was told that words would be shown and that his task was to say aloud, as quickly as possible, the color of ink in which each word was printed. Several examples were shown during the instructions, to familiarize *S* with the procedure and with the four acceptable color names. Ten practice stimuli were then presented, followed by the 128 experimental stimuli. The stimulus duration was 1 sec., and the interstimulus interval approximately 4 sec. A vocal warning ("ready") was given just prior to the onset of each stimulus. Response latencies were measured using a .01-sec. stop clock triggered by a voice key. Latencies and errors were recorded by *E*.

---

[1] This study was done as a Laboratory in Learning class project. Thanks are due to Steven Keele for comments on the manuscript.

[2] Requests for reprints should be sent to Douglas L. Hintzman, Department of Psychology, University of Oregon, Eugene, Oregon 97403.

Results and discussion.—Only correct response latencies are presented. By conditions, the percentages of incorrect responses (including "false starts") were: S, 1.1; D, 13.2; N, 1.9; and C, 2.8. As would be expected, the classical Stroop condition produced the highest error rate.

Every *S*'s mean correct response latency was computed for each condition in each of the four blocks, and these means were submitted to a 4 × 4 repeated-measures analysis of variance. Stimulus conditions had a significant overall effect, $F\ (3, 51) = 34.23$, $p < .001$, and there was a slight decrease in latencies over blocks, $F\ (3, 51) = 4.37$, $p < .01$. The Conditions × Blocks interaction was marginally significant, reflecting a tendency for Cond. D to improve with practice more than the others, $F\ (9, 153) = 2.41$, $p < .05$.

Overall mean latencies by conditions (in msec.) were: S, 653; D, 799; N, 694; C, 696. Planned comparisons showed that Cond. S was significantly faster than Cond. C, $t\ (17) = 3.08$, $p < .01$, and Cond. D significantly slower than C, $t\ (17) = 6.15$, $p < .001$. Conditions N and C did not differ significantly. This lack of a difference between Cond. N and C is consistent with Keele's (1972) results, suggesting that the effect of neutral words on color naming found by Klein (1964) is eliminated when the discrete-trial procedure is used.

The facilitation found in Cond. S tends to confirm Keele's (1972) claim that selective attention operates after memory retrieval. If the classic Stroop condition were due primarily to interference with the ability to encode ink color, then it should make little difference whether or not a color word matches the color of ink; interference should be produced in either case. The present results indicate that a compatible color word enhances color-naming performance not only in comparison with the classic Stroop condition, but also in comparison with the neutral control. This is what would be expected if the Stroop effect were due to covert response competition. Of course, one cannot be certain that color words produced *no* encoding interference in the present experiment. However, if such interference was present, its effect must have been relatively small. The present findings argue that the Stroop effect is primarily an output, rather than an input phenomenon.

REFERENCES

HOCK, H. S., & EGETH, H. Verbal interference with encoding in a perceptual classification task. *Journal of Experimental Psychology*, 1970, **83**, 299–303.

KEELE, S. W. Attention demands of memory retrieval. *Journal of Experimental Psychology*, 1972, **93**, 245–248.

KLEIN, G. S. Semantic power measured through the interference of words with color-naming. *American Journal of Psychology*, 1964, **77**, 576–588.

PRITCHATT, D. An investigation of some of the underlying associative verbal processes of the Stroop color effect. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 351–359.

(Received September 23, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 460–462

ASSOCIATIVE REACTION TIME, MEANINGFULNESS, AND MODE OF STUDY IN FREE RECALL[1]

DAVID LOCASCIO[2] AND RONALD LEY[3]
State University of New York at Albany

Eighty *S*s were tested for free recall of three lists of 12 CVCVCs each, all lists equally balanced with respect to short- and long-latency (SL and LL) associative reaction time (RT) units within levels of high and low meaningfulness (HM and LM). Four groups of 20 *S*s each were instructed to study the lists under one of four different study modes. The *S*s instructed to make either single associations (SA) to the units during study or multiple associations (MA) or to rehearse silently (SR) recalled more units than *S*s instructed to repeat the units aloud (RP). SL-RT units were recalled more frequently than LL-RT units for the SR and RP groups. HM units were recalled more frequently than low-M units for all groups. The results were interpreted as support for the hypothesis that making associations to verbal units facilitates recall and that RT is an index of the facility with which associations can be made.

Ley and Locascio (1970b) have proposed a modification of the two-stage analysis of paired-associate learning (PAL) wherein recognition and recall are viewed as indicants of two separate memory processes rather than as two indicants of a single memory process. In this modification it is hypothesized that recognition is dependent on a perceptual encoding response, the stability of which is indexed by the meaningfulness (M) of the verbal unit (a formulation consistent with that of Martin, 1968); and recall is hypothesized to be dependent on an association response to the verbal unit, the strength of which is indexed by associative reaction time (RT), the time elapsing between the presentation of a verbal unit and the first association produced. This formulation is an outgrowth of a series of PAL studies which have shown that (*a*) RT of response terms (the terms which must be recalled) is related to performance (Ley & Locascio, 1970a), (*b*) RT of stimulus terms (the terms which need only be recognized) is not related to performance (Ley & Locascio, 1970c), and (*c*) in backward recall following PAL, short-latency RT stimulus terms are *recalled* more frequently than long-latency RT terms (Ley & Locascio, 1970b). With respect to the effect of M, the findings of Ley and Locascio are consistent with the hypothesis that both the stimulus and response terms of PAs must be recognized before recall of either can occur; i.e., in both forward and backward PAL, M is related to performance whether the M-indexed verbal unit is in the response or stimulus positions of a PA list.

The primary purpose of the present study was to extend the investigation of the proposed relationship between RT and recall from PAL to free recall. If RT is related to recall performance in general (i.e., if the effect of RT is not an artifact of PAL), then in a free-recall task it would be expected that within levels of equated M, short-latency RT units will be recalled more frequently than long-latency RT units.

The secondary purpose of the present study was to test the hypothesis that in a free recall task, *S*s instructed to make associations to verbal units during study will recall more units than *S*s instructed to repeat the units aloud, as in the typical familiarization procedure. This prediction stems from a basic assumption of the Ley and Locascio (1970b) two-process modification, viz., that making an association to a verbal unit during study facilitates its recall, and that RT is related to recall in that it indexes the facility with which an association can be made. If this assumption regarding the relationship between making associations to verbal units and facility of recall of the units obtains, then free recall should be a function of the opportunity *S*s have to make associations to verbal units during study. Additionally, if RT indexes the associative strength of the initial response, this response should be most important in supporting recall, and additional associations should have little, if any, facilitative effect on recall.

Method.—The *S*s were 80 undergraduate, paid volunteers (28 males and 52 females) from the State University of New York at Albany, naive with respect to prior experience in verbal learning studies.

The materials were 36 CVCVC verbal units selected from the 319 units assessed for scaled-rated M value (m') by Locascio and Ley (1972) and assessed previously for RT by Taylor and Kimble (1967). Using m' as the primary criterion, 18 CVCVCs were selected which had short-latency (SL) RT values: 9 from the high-meaningfulness (HM) end of the m' distribution and 9 from the low-meaningfulness (LM) end of the m' distribution; and 18 CVCVC verbal units were selected which had long-latency (LL) RT values: 9 from the HM end and 9 from the LM end of the m' distribution.

All verbal units with an m' value of 2.00 or less were chosen as representative of LM. The RT values of this LM sample were then taken from Taylor and Kimble's (1967) list. The mean and standard deviation of these RT values were calculated. Verbal units which fell one or more standard deviations below the mean were considered for the purposes of the present study to be SL-LM units, and verbal units which fell one or more standard deviations above the mean were considered to be LL-LM units. Nine CVCVC verbal units were then selected from each of these RT groups. The same procedure was used in selecting the HM-SL and HM-LL verbal units. The *m′* and RT values for each of the 36 CVCVCs selected are given in Table 1, from which it can be seen that the mean *m′* values of the SL- and LL-RT verbal units are essentially the same within both HM and LM levels.

---

[1] This research was supported in part by a State University of New York Research Foundation Grant in Aid to the second author.

[2] Now at Fairleigh Dickinson University, Madison, New Jersey.

[3] Requests for reprints should be sent to Ronald Ley, State University of New York at Albany, Albany, N.Y. 12203.

The 36 CVCVCs were assigned to three lists, each list containing three randomly assigned CVCVC verbal units from each of the four classifications (HM-SL, HM-LL, LM-SL, LM-LL). The lists of CVCVCs were then typed in random order on a paper tape for presentation on a Lafayette memory drum.

The order of the CVCVCs in each of the three lists was determined randomly with the following qualifications: (*a*) Each classification group (HM-SL, HM-LL, LM-SL, LM-LL) was represented by one verbal unit in each third of the list. (*b*) No two verbal units from the same classification group appeared immediately adjacent to one another on any list. (*c*) The position that the verbal unit from each classification group occupied in each third of the list was different from list to list; e.g., if the HM-SL unit appeared second in the first third of List 1, then the unit that was HM-SL on List 2 could not occupy that position on the first third of List 2, etc. (*d*) The order of list presentation for each *S* was balanced so that each of the three lists was presented first, or second, or third an approximately equal number of times.

Two practice lists, containing 12 different CVCVC verbal units each, were constructed in the same manner as the three test lists and typed on a paper tape. The practice lists were similar to the test list, with the verbal units selected from the remaining 283 CVCVC units assessed for *m′* and RT. Formal similarity between the verbal units on the practice lists and those on the test lists was avoided.

The *S*s were instructed to learn 36 CVCVC verbal units in a free-recall task under one of four modes of study: Repeated Pronunciation (RP), Single Association (SA), Multiple Association (MA), and Silent Rehearsal (SR), with an equal number of *S*s ($n = 20$) randomly assigned to each of the four study-mode treatment groups.

TABLE 1

MEAN ASSOCIATIVE REACTION TIME (RT), MEAN RATED ASSOCIATIVE FREQUENCY (*a′*), AND MEAN SCALED RATED ASSOCIATIVE FREQUENCY (*m′*) FOR 36 CVCVC VERBAL UNITS EMPLOYED IN A FREE-RECALL LEARNING TASK

RT	CVCVC	RT	*a′*	*m′*
High meaningfulness				
Short latency	TENOR	1.88	4.05	3.86
	DIGIT	1.80	4.00	3.80
	RAVEN	1.78	4.00	3.80
	TOTEM	1.84	3.58	3.26
	DOZEN	1.81	4.26	4.16
	PATER	1.90	3.20	2.87
	LILAC	1.91	4.10	3.92
	DIVAN	1.86	2.94	2.55
	ROBIN	1.73	4.12	3.95
Long latency	FAVOR	2.80	3.94	3.70
	BASIS	2.52	3.72	3.42
	HABIT	2.70	3.94	3.70
	RADAR	2.55	4.08	3.90
	RUMOR	2.80	4.23	4.12
	LAGER	2.68	3.54	3.23
	VICAR	2.82	3.26	2.92
	CYNIC	3.35	3.56	3.25
	BIGOT	3.42	3.90	3.65
Low meaningfulness				
Short latency	JUREL	2.39	1.72	1.09
	GORAL	2.62	1.64	.97
	RATAL	2.64	1.66	1.00
	LIMEN	2.40	1.69	1.05
	NEROL	2.77	1.41	.52
	SALOL	2.70	1.54	.82
	TONUS	2.29	1.69	1.05
	BOLUS	2.78	1.66	1.00
	XYDER	2.48	1.66	1.00
Long latency	ZUREN	4.65	1.48	.70
	XABIN	3.86	1.87	1.27
	MENAD	3.72	1.64	.97
	VUTAW	4.26	1.36	.36
	HYRAX	4.04	1.66	1.00
	NIDUS	3.71	1.62	.94
	JEHAD	3.85	1.66	1.00
	POLEF	3.74	1.84	1.23
	TAROP	3.99	1.56	.85

All *S*s were read standard instructions appropriate to the assigned study-mode group, which in essence explained the mode of study to be followed and the recall procedure that would be employed. The practice lists were then presented to *S* at a 4-sec. rate, followed by a 90-sec. period during which the verbal units were to be orally recalled. Immediately following the recall period for Practice List 2, the three test lists were presented at a 4-sec. rate, followed by a 90-sec. recall period. A 20-sec. interval was employed between the end of one recall period and the presentation of the subsequent test list.

In the RP mode of study, *S* was instructed to pronounce repeatedly each unit aloud as a "word" during the 4-sec. presentation period. In the SA mode of study, *S* was instructed to make a free association response to each unit and then repeat aloud this initial free associate during the 4-sec. presentation period. In the MA mode of study, *S* was instructed to make aloud as many different free-association responses as he could to each unit during the 4-sec. presentation period. In the SR mode of study, *S* was instructed to study silently each verbal unit during the 4-sec. presentation period.

The learning score for *S* consisted of the number of CVCVCs recalled within each of the four RT-M categories; thus scores could vary from 0 to 9. The design employed was a three-variable mixed factorial, in which mode of study (RP, SA, MA, SR) was a between-*S*s factor, and RT (SL and LL) and *m′* (HM and LM) were within-*S*s factors.

Results.—The mean numbers of SL- and LL-RT CVCVCs recalled under each of the four modes of

TABLE 1

ORIGINAL LEARNING: PROPORTION OF SUBJECTS WITHIN GROUP AND LIST CONDITIONS

Cond.	Trials-to-criterion intervals				
	0–5	6–10	11–15	16–20	21–25
Experimental					
List A-B	.20	.20	.40	.15	.05
List B-A	.05	.30	.35	.20	.10
Control					
List A-B	.10	.50	.30	—	.10
List B-A	.10	.40	.30	.20	—

ences were found between the means of the experimental and control groups (12.25 and 10.85, respectively), the data suggest that the two groups were comparable.

Test-trial scores were also analyzed by means of a 2 × 2 analysis of variance. This analysis indicated that the performance between the two groups was significantly different, $F\ (1, 56) = 8.335$, $p < .01$, and that the performance of *S*s on the two PA lists was not significantly different, $F < 1$. Also, the interaction effects were not found to be significant, $F < 1$. An examination of the mean number of correct anticipations made by the experimental and control *S*s (5.6 and 6.6, respectively) indicated that more forward associations were established during learning than backward associations. Also, since neither analysis revealed any list differences, it is suggested that the associative stage in learning A-B and B-A pairs was not intrinsically different in difficulty and that the difference in test-trial scores does not reflect the confounding of pairing direction and associative directionality.

Although the nature of the pairs in this experiment provided for maximal backward associations, forward associations were still found to be superior to backward associations. This finding is in direct contradiction to Houston's (1964) reported finding of no difference between forward and backward associations. However, a close examination of the procedure on the test trial in Houston's experiment suggests that his measure of the strength of the directions of associations might not have been dependent solely on the direction being tested. He used a 5:2 sec. rate of presentation on the test trial to allow for differences in the amount of material to be recalled. However, Ekstrand (1966) has cautioned that the test for backward associations should proceed at a fairly rapid rate to reduce the possibility of *S*s covertly running through the pairs in the forward direction until the correct stimulus term is located. With a seven-pair list and a 5-sec. stimulus presentation on the test trial, *S*s tested for backward associations could have run through a good part, if not all, of the learned list before the anticipation period ended. Therefore, the number of backward associations reported by Houston might contain responses which *S*s made on the basis of the learned forward associations. The present experiment did not afford *S*s as much opportunity to supply the correct backward associations on the basis of learned forward associations.

The results of the present experiment clearly indicate the superiority of forward associations. However, this finding, in and of itself, may not necessarily be evidence against the principle of associative symmetry. The results must be interpreted in terms of the methodological considerations outlined above. In other words, it must be concluded that although effective functional stimuli and equally difficult A-B and B-A pairs may be necessary conditions, they do not appear to be sufficient conditions for the demonstration of associative symmetry. Other conditions which might also be necessary for the demonstration of associative symmetry, and to which the present experimental design was not sensitive, are the analysis of backward and forward associations within pairs and the use of a learning method and materials which would demand equal attention to both the stimulus and response terms and which would equate the learning and test situations along a number of dimensions (e.g., response set and item availability).

REFERENCES

ASCH, S. E., & EBENHOLTZ, S. M. The principle of associative symmetry. *Proceedings of the American Philosophical Society*, 1962, **106**, 135–163.

EKSTRAND, B. R. Backward associations. *Psychological Bulletin*, 1966, **65**, 50–64.

HOUSTON, J. P. S–R stimulus selection and strength of R–S associations. *Journal of Experimental Psychology*, 1964, **68**, 563–566.

UNDERWOOD, B. J., HAM, M., & EKSTRAND, B. R. Cue selection in paired-associate learning. *Journal of Experimental Psychology*, 1962, **76**, 470–474.

WOLLEN, K. A. Effects of maximizing availability and minimizing rehearsal upon associative symmetry in two modalities. *Journal of Experimental Psychology*, 1968, **77**, 626–630.

(Received October 18, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 465–467

REHEARSAL OF INDIVIDUAL ITEMS IN SHORT-TERM MEMORY

GARY F. MEUNIER,[1] DOUGLAS RITZ,[2] AND JO ANN MEUNIER

Ball State University

An experiment was performed in an attempt to examine the function of rehearsal in the retention of individual verbal items. Varying amounts of unrestricted rehearsal and rehearsal interference were manipulated between *S*s. Recall was taken after each item and again after all items had been presented. The *S*s were not informed of the final recall task until all items had been presented. The wide difference in results on the two recall tests for *S*s given extra rehearsal time suggested that *S*s utilized rehearsal primarily to maintain items in short-term memory.

Much recent work in the psychology of immediate memory has derived from a report published by Peterson and Peterson (1959). These investigators found that memory for briefly presented individual verbal units decreased steadily over intervals from 3 to 18 sec. if *S* was not allowed to rehearse the unit. A great many replications of this work attest to the stability of this result (e.g., Murdock, 1961).

An aid to the understanding of processes involved in this type of retention task has been proposed by Waugh and Norman (1965). These investigators suggested that an item is held in a short-term storage of relatively limited capacity (short-term memory, STM) and encoded into a longer term storage of much greater capacity (long-term memory, LTM) by a process of rehearsal. The longer an item is held in STM the greater the probability that it will be encoded into LTM. And an item is "forgotten" from STM when rehearsal is inhibited.

Implications of this conceptualization are generally useful in the interpretation of experimental results from studies directly concerned with STM phenomena. For example, presentation of an interfering task immediately after the stimulus material in the Peterson and Peterson (1959) paradigm produces a very high rate of forgetting which is inversely related to the duration of the interfering task (e.g., Murdock, 1961; Peterson & Peterson, 1959). The high rate of forgetting is presumably due to the inability of *S* to retain the items in STM or encode them into LTM. When a period of unrestricted rehearsal, on the other hand, is interjected between stimulus presentation and recall, forgetting is typically delayed. This "inhibition of forgetting" is related directly to the amount of rehearsal provided (Stanners & Meunier, 1969). A slight confusion in interpretation is present, however, in experiments in which extra rehearsal is allowed. This difficulty in interpretation stems from the fact that rehearsal is postulated to have the dual function of retaining items in STM and encoding them into LTM. At the moment of recall, thus, items may be stored in LTM, STM, or some combination of the two.

The present study was an attempt to gain further purchase on the storage of items given unrestricted rehearsal. Specifically, using a modification of the Peterson and Peterson (1959) task, retention of items given various amounts of rehearsal time was tested after each item had been presented and again after all items had been presented. While performance for *S*s given extra rehearsal was ambiguous on the first test (due to the fact that *S*s could be recalling items from either store), second-test performance was expected to be a relatively "pure" measure of LTM retention and, hence, give important clues as to the storage of items recalled on the first test.

Method.—Fifty students from introductory psychology classes at Ball State University served as *S*s. All *S*s were given a slight addition to their final course grades as an inducement to participate.

Stimulus materials consisted of 25 low-association-value (29%) trigrams selected from the Witmer (1935) scaling. In addition, five common three-letter words were selected for use as practice items. Thirty different three-digit numbers were also selected for use in a backwards counting task.

The type of activity which followed the presentation of an item (counting backwards from a three-digit number or unrestricted rehearsal) was manipulated between *S*s and the length of the retention interval (0, 3, 6, 9, and 18 sec.) was manipulated within *S*s. Twenty-five *S*s served in the counting (C) and no-counting (NC) groups, and each *S* was presented with 25 trigrams, 5 at each of the retention interval lengths.

Stimuli were presented auditorially by a Webcor tape recorder to individual *S*s. Individual letters and digits were recorded in a monotone voice by a male graduate student in psychology. Total presentation time of a complete trigram or number was 1 sec., and a three-digit number followed each of the trigrams by approximately .5 sec. The *S*s in the NC condition were told to ignore the number; *S*s in the C condition were instructed to count backwards aloud from it by three's until the recall signal appeared. The *S*'s verbal recall was signaled by a red light and buzzer mounted on a panel which shielded *S* from *E* and the programming equipment. A 30-sec. period was allowed between onset of the recall signal and *E*'s verbal warning that the next

---

[1] Requests for reprints should be sent to Gary F. Meunier, Department of General and Experimental Psychology, Ball State University, Muncie, Indiana 47306.

[2] An earlier version of this experiment comprised a Master of Science thesis by the second author.

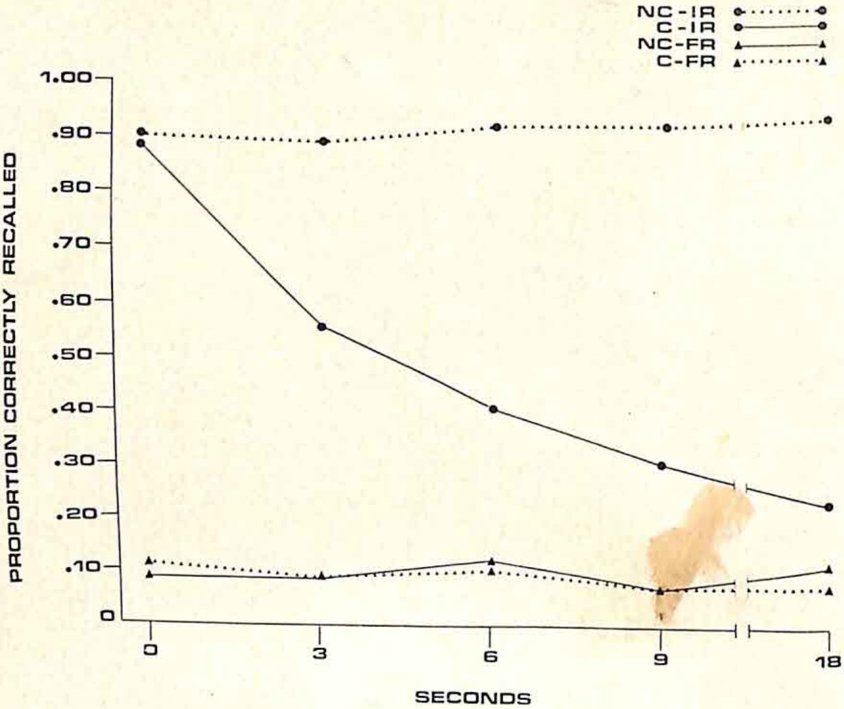


FIG. 1. Proportion of items recalled as a function of experimental condition and length of retention interval.

trial was about to begin. After all trigrams had been presented, *S* was given an $8\frac{1}{2} \times 11$ in. sheet of blank paper and asked to recall any trigrams he could. The *S* had not been given previous knowledge of the final recall task. A total of 5 min. was allocated for the final recall.

Five practice trials were given each *S* at intervals of 2, 4, 7, 14, and 20 sec. Experimental procedures during the practice trials were, with the exception of a final recall, indistinguishable from actual experimental trials. Counterbalancing procedures required that each trigram was followed by each retention interval length an equal number of times in both C and NC groups. In addition, each trigram and interval length appeared equally often in each serial presentation position for both groups.

Results.—Due to the unique nature of the data, separate analyses of variance were conducted on initial recall (IR) and final recall (FR) performance. The experimental unit for all statistical analyses was a whole trigram and the .05 level was arbitrarily selected as the minimum requirement for statistical significance.

Results from the analysis of IR data strongly support the graphical representation of Fig. 1. Both main effects of counting or not counting (C–NC) and intervals (I) were significant as was the C–NC × I interaction. Respective *F* values were: C–NC, $F(1, 48) = 271.69$, $p < .001$; I, $F(4, 192) = 17.14$, $p < .001$; and C–NC × I, $F(4, 192) = 28.22$, $p < .001$. Recall was obviously inhibited by the backwards counting task and was unaffected by extra rehearsal time. Thus, additions to rehearsal beyond a certain minimum seemed to have little effect on overall level of retention but the interference task seriously inhibited retention. Further evidence for this conclusion was provided by separate analyses of variance performed on the C and NC data. The main effect of intervals was significant only in the analysis of the C data, $F(4, 96) = 51.82$, $p < .001$.

Analysis of the FR data confirmed the quite surprising results suggested by the FR lines in Fig. 1; no effects in the analysis approached significance. This finding is surprising in light of the large difference in amount of rehearsal time given different *S*s and the large difference in retention for different groups on the IR task. Although the opportunity for rehearsal was a strong predictor of IR data, there would seem to be no relationship between retention on the FR task and rehearsal time.

Discussion.—Analysis of initial recall performance seems quite straightforward: Retention is nearly perfect when *S* is allowed unrestricted rehearsal and declines directly with the amount of rehearsal interference provided. These results seem generally supportive of suggestions from the Waugh and Norman (1965) conceptualization that (*a*) rehearsal interference allows neither the maintenance of items in STM nor the encoding of items into LTM, and (*b*) unrestricted rehearsal allows both.

As mentioned earlier, one difficulty with these data is that they do not allow the determination of where the items were stored when recall was signaled.

Fortunately, analysis of final recall data does give a clue as to how these items were retained. This clue derives from the fact that items recalled during the final recall period must have been retrieved from LTM. The fact that final recall performance was quite low (below 15%) for all *S*s suggests that the extra rehearsal time given NC *S*s was not utilized for the encoding of items into LTM. Hence, all *S*s must have retrieved items for the initial recall test from STM and utilized the extra rehearsal time primarily for the maintenance of items in STM only.

These conclusions regarding the use of unrestricted rehearsal time are surprising in light of the suggestion from the Waugh and Norman (1965) model that the probability of an item being encoded into LTM is a direct function of the amount of rehearsal provided. Although other explanations are possible (e.g., items were encoded into LTM but forgotten before final recall), the present conceptualization seems most parsimonious.

REFERENCES

Murdock, B. B., Jr. The retention of individual items. *Journal of Experimental Psychology*, 1961, **62**, 618–625.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

Stanners, R. F., & Meunier, G. F. Pronunciability and rehearsal time in short-term memory with controlled acquisition. *Journal of Experimental Psychology*, 1969, **80**, 359–365.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89–104.

Witmer, L. R. The association-value of three-place consonant syllables. *Journal of Genetic Psychology*, 1935, **47**, 337–360.

(Received October 27, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 467–469

EFFECTS OF HIGH-PRIORITY EVENTS ON RECOGNITION OF ADJACENT ITEMS

LYNN S. SCHULZ[1] AND RONALD B. STRAUB

University of Delaware

High-priority events (HPEs) were names of famous people inserted in 15-item lists of common words. The following five conditions were examined with four lists per condition: a single HPE in Input Position 7; two HPEs in Positions 7 and 8; two HPEs in Positions 7 and 9; two HPEs in Positions 7 and 12; and control lists with no HPEs. Significant decrements were observed in recognition of items immediately preceding and immediately following HPEs. Detrimental effects were larger for preceding items when only a single HPE occurred in a list; with two HPEs present, preceding and following items were affected to an equal degree. The results confirm previous findings with famous-name HPEs and indicate that the effects are independent of type of recognition task employed.

The insertion of a high-priority event (HPE) in a list of unrelated items depresses recognition of the immediately preceding and the immediately following items when the HPEs are names of famous people (Schulz, 1971). When photographs of nudes are used as HPEs in lists of line drawings, no detrimental effects are observed on immediately preceding items, but profound anterograde effects result which persist over several items following each HPE (Ellis, Detterman, Runcie, McCarver, & Craig, 1971). Discrepant results in the two studies might stem from variations in stimuli, list length, instructions, or type of recognition test: forced-choice (Schulz, 1971) versus yes–no measures (Ellis et al., 1971). Possible importance of the type of task employed is suggested by the differential effects of HPEs in recall, where retrograde losses predominate (Saufley & Winograd, 1970; Tulving, 1969), and in forced-choice recognition where bilateral effects are observed (Schulz, 1971).

The present study was designed to assess the effects of famous-name HPEs on the recognition of surrounding items as measured by a yes–no recognition test. Conditions were included to determine the effects of two HPEs occurring at various separations in the same list since earlier data had provided somewhat inconclusive results using forced-choice procedures (Schulz, 1971).

Method.—The *S*s were 90 undergraduate students from the University of Delaware who participated in the experiment to fulfill a course requirement. No *S* had previous experience with the recognition task.

Testing was accomplished in large group sessions with all *S*s viewing the same 21 lists of 15 items each; order of presentation remained constant for all groups. List items were projected onto a large screen with a Dunning Animatic filmstrip projector. A Lafayette repeat-cycle timer was used to control

---

[1] Requests for reprints should be sent to Lynn S. Schulz, College of Education, University of Delaware, Newark, Delaware 19711.

TABLE 1

Percentage of Items Recognized at each of the Fifteen Input Positions

List	Input position														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Control	93	72	82	76	67	(76)	69	(70)	(72)	(79)	(75)	73	(75)	88	92
HPE 7	82	77	83	74	73	(55)	[95]	(58)	78	68	66	79	76	89	89
HPE 7, 8	90	75	71	69	65	(61)	[89]	[96]	(58)	70	72	66	73	85	93
HPE 7, 9	86	83	65	74	69	(57)	[96]	(51)	[92]	(57)	68	62	66	80	84
HPE 7, 12	90	87	79	66	67	(60)	[94]	(61)	63	66	(54)	[93]	(51)	86	77

Note.—Percentages are based on 360 observations (Lists × *S*s). Brackets indicate locations of HPEs; parentheses surround IPIs, IFIs, and corresponding control values.

presentation at a 1-sec. rate (not including minimal change-over interval).

The first list presented contained no HPEs and was administered for practice purposes only. The remaining 20 lists represented the following five conditions: four lists had an HPE at Input Position 7 (HPE 7); four lists contained two HPEs at Positions 7 and 8 (HPE 7, 8); four lists had HPEs at Positions 7 and 9 (HPE 7, 9); four lists had HPEs at Positions 7 and 12 (HPE 7, 12); and four lists had no HPEs (Control). Presentation order of the lists was determined randomly except that no condition was permitted to occur more than twice in succession. Three asterisks followed each list as a signal to begin the recognition test.

Names of famous people were used as HPEs (e.g., Kennedy, Einstein); all other list items were common words with Thorndike-Lorge frequencies greater than nine per million. To construct the recognition tests, distractor items were chosen to match each list item in terms of Thorndike-Lorge frequency, initial letter, and number of syllables. Distractors for HPEs were famous names with the same initial letter and the same approximate length. Distractors and list items were randomly arranged in a column of 30 words with the restriction that HPEs and their distractors were always tested last (i.e., the bottom of the column). The *S*s were to circle Y (yes) or N (no) printed next to each item depending on whether it had or had not occurred in the list they just viewed.

All *S*s were instructed that some of the lists contained one or more names of famous people, and that it was very important that all famous names be recognized correctly. It was stressed that recognition decisions were to be made in sequence, without leaving any items unanswered, and without returning to difficult decisions. When all *S*s had completed a recognition test (which took approximately 30 sec.), the next list of words was presented on the screen.

Results and discussion.—Table 1 shows the percentages of items recognized at each input position for each of the five conditions. Brackets enclose data from positions at which HPEs occurred; parentheses indicate the immediately preceding items (IPIs), the immediately following items (IFIs), and the comparable control values.

Recognition of HPEs averaged 94% over all conditions, despite the fact that these items were not tested until the last positions in the recognition task. Overall list recognition averaged 75%; differences between conditions in overall recognition were not statistically significant.

Analyses of variance were performed to make separate comparisons of each HPE condition to control performance. The basic data were numbers of recognition errors in each list (collapsed across *S*s) on IPIs, IFIs, and control items in comparable input positions. In each analysis, conditions represented a between-lists variable with two levels (HPE vs. Control) and position represented a within-lists variable, the number of levels being dependent on the number and spacing of HPEs in that condition. For example, data from four positions (6, 8, 10, 13) were considered in analyzing the HPE 7, 12 condition, three positions were relevant for HPE 7, 9; and two positions for both HPE 7 and HPE 7, 8 conditions.

All analyses resulted in highly significant main effects of condition, indicating that fewer adjacent items were recognized in each HPE condition as compared to the control, Fs $(1, 6) = 62.61$, 11.36, 39.47, 50.71 for HPE 7; 7, 8; 7,9; and 7, 12, respectively, ps $< .05$. The effect of position did not approach statistical significance in any of the analyses, but a Condition × Position interaction was observed in comparison of HPE 7 and Control conditions, F $(1, 6) = 27.53$, $p < .01$. This interaction reflected greater decrements for IPIs than for IFIs in the HPE 7 condition. It should be noted that the magnitude of the interaction effect results in part from opposite directional tendencies in the control group. No interactions were found in analyses of other conditions, Fs < 1.

Earlier findings of detrimental effects of famous-name HPEs on the recognition of adjacent items are confirmed by the present data, and the generality of the phenomenon is extended to yes–no measures of recognition. Predominance of retrograde effects in the HPE 7 condition can be related to previous observations of larger recognition decrements for IPIs than IFIs when only single HPEs occur in a list (Schulz, 1971, Exp. I). With two HPEs present, prior evidence indicated significant effects on all adjacent items in an HPE 7, 9 condition; in HPE 7, 8 and 7, 12 conditions, however, the only salient effects in forced-choice recognition were decrements on IPIs (Schulz, 1971, Exp. III). The present study shows more conclusively that IPIs and IFIs are affected to an equal degree when two HPEs occur in a list.

The consistent findings of impaired recognition of IPIs and IFIs may be interpreted as a result of loss of effective presentation time for both kinds of items. The onset of a famous-name HPE is presumed to terminate normally ongoing consolidation or encoding of the IPI; rehearsal of the HPE just after its presentation is assumed to detract from effective presentation time of the IFI.

Type of recognition task employed is ruled out by the present study as an explanation for the differential effects of famous names and nudes as HPEs. The possibility remains that line drawings may require less encoding or processing time than verbal items; IPIs of the former type might therefore remain unaffected by HPEs except at very rapid rates of presentation. The severe anterograde effects could be attributed to a failure to attend to the line drawings after the appearance of nude pictures. Further research should focus on such characteristics of the stimulus materials as factors potentially responsible for the discrepant results.

REFERENCES

Ellis, N. R., Detterman, D. K., Runcie, D., McCarver, R. B., & Craig, E. M. Amnesic effects in short-term memory. *Journal of Experimental Psychology*, 1971, **89**, 357–361.

Saufley, W. H., & Winograd, E. Retrograde amnesia and priority instructions in free recall. *Journal of Experimental Psychology*, 1970, **85**, 150–152.

Schulz, L. S. Effects of high-priority events on recall and recognition of other events. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 321–330.

Tulving, E. Retrograde amnesia in free recall. *Science*, 1969, **164**, 88–90.

(Received November 1, 1971)

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 469–471

PREFERENCE FOR INFORMATION ABOUT AN UNMODIFIABLE BUT REWARDING OUTCOME[1]

JOHN J. FUREDY[2] AND FELIX KLAJNER

University of Toronto

Postexperimental questionnaires assessed preference for signaled (5-sec. signal with onset preceding outcome onset by 7.5 sec.) over unsignaled rewarding outcomes, where the outcomes were 15-sec. cooling air puffs to 40 *Ss*' faces in a continuously hot (110°F) and humid (80% relative humidity) environment. Signals allowed no outcome modification, yet there was a highly reliable preference for signaled puffs, and this preference was independent of sex of *Ss*, willingness to perform an easy task for the puffs, and score on a postexperimentally administered achievement scale. The contrast between these clear signaled-preference results, and previous difficulties in obtaining reliable preferences for signaled unmodifiable noxious (shock) outcomes is discussed.

It is both commonly believed, and is consistent with current theoretical formulations (e.g., Perkins, 1968), that organisms prefer to obtain information about noxious outcomes even when the information does not permit modification of those outcomes. The evidence for this notion comes mainly from the occurrence of the preference-for-signaled-shock (PSS) phenomenon, which has been reported both in (*a*) animals (e.g., Perkins, 1968) and (*b*) humans (Lanzetta & Driscoll, 1966). However, (*a*) as argued by Furedy and Doob (1972), the animal PSS phenomenon may have occurred only because the shocks were modifiable (cf. Biederman & Furedy, 1970, for indirect evidence to support this view), while (*b*) the human PSS effect was not found in a number of studies by Furedy and Doob (1972), who, in discussing their PSS results, argued for the conclusion that Lanzetta and Driscoll's (1966) reference experiment in favor of the PSS phenomenon "was not optimally designed to test for a preference for signaled over unsignaled shock [Furedy & Doob, 1971, p. 8]." At this stage, then, although the issue is far from finally resolved, it is at least doubtful whether there really is a preference for information about unmodifiable noxious outcomes.

On the other hand, there is both less doubt about and less evidence concerning preference for information about unmodifiable rewarding outcomes. In a hot and humid environment, a biologically significant outcome of this sort is provided by a puff of cooling air to *S*'s face, and forms of this type of outcome have been shown to be positively reinforcing for humans (e.g., Furedy, 1967; Paul, Eriksen, &

---

[1] This research was supported by grants from Canada Council (S70-1710) and from the National Research Council of Canada (APA 222) to John J. Furedy.

[2] Requests for reprints should be sent to John J. Furedy, Department of Psychology, University of Toronto, Toronto 181, Ontario, Canada.

Humphreys, 1962). The experiment to be reported here used a rewarding-puff outcome, and evidence was sought concerning two issues: (*a*) whether the preference-for-signaled-puff (PSP) phenomenon occurs reliably as a group effect, and (*b*) if so, whether this sort of preference for information about an unmodified but rewarding outcome is related to other variables such as the sex of the *S*s.

Method.—The *S*s were 20 male and 20 female undergraduates from the University of Toronto who were paid $1.50 for participating.

The *S* was comfortably seated before a plywood water-resistant hot box and his face was loosely strapped into a padded oval opening in the front panel of the box. The dimensions of the internal part of the box visible to *S* were 46 cm. (height), 61 cm. (width), and 58 cm. (depth). Internal-box temperature was thermostatically controlled within ±3° F. at 110°; relative humidity was 80%, ±5%, generated by two vaporizers hidden from *S*'s view, and controlled by a humidistat. The 15-sec. puff was delivered through a 4.4-cm.-diam. metal pipe, the opening of which was 28 cm. away from *S*'s face at an upward angle of 40°. Puff onset and offset were controlled by a solenoid valve, and the source of the puff was a continuously moving flow of air produced by a fan rotating at 3,060 rpm., driven by two Redmond motors (Model AJ2H01AY). The air was drawn from the air-conditioned room containing *S* and the box, but, because of the great heat transfer, the puff received by *S* was close to the 110° F. internal box temperature. However, because of the high humidity of the internal-box environment, the puff, through its evaporative function, felt pleasantly cool. Evidence that heat transfer renders the puff temperature at origin of little importance comes from a pilot study where the puff was sometimes delivered directly from an air-conditioning unit (at temperatures less than 45° F.) rather than from the room air; *S*s could not tell the difference between the two puffs. The signal for the puff consisted of an increase in general illumination in the box from 2.3 to 7.7 ftl.

Two other features visible to *S* and related to a task *S* sometimes had to perform, were: (*a*) two 28-v. lights (red or blue) mounted in front of *S* on the back wall of the chamber, and (*b*) between the two lights, a white semi-circular 14-cm.-diam. plastic panel, which, when lit up, showed the word "choice." Finally, for the task, there were four pushbutton momentary-action switches mounted on the outside of the box on a foam-padded panel on which *S*'s right arm rested, and depression of each switch resulted in *S* receiving, through earphones, a 1.5-sec., 69-db. tone of 350, 1,200, 2,000, or 3,500 Hz. The *E*, who was in a separate room, was in constant contact with *S* via *S*'s earphones and a microphone strapped to *S*'s throat.

Before his face was strapped in, *S* was shown the four task switches, and told that these "keys" would later be referred to as "one," "two," "three," and "four" from left to right. The task, as instructed to *S* through the intercom, was to press these keys in the sequence 4-2-3-1, and *S* was told that this work would sometimes be necessary for receipt of each puff. Such "work" and "no-work" periods were differentially indicated by the onset of the red and blue lights (counterbalanced between *S*s); during no-work periods, *S*s were to get the puffs without having to perform the task. The *S*'s face was exposed to the heated-box environment for approximately 5 min. before the beginning of the experiment.

In Stages 1 and 2 of the experiment, each *S* received a work period followed by a no-work period, or vice versa. Each period comprised six 60-sec. trials with no interval between trials. Each 60-sec. trial consisted of: (*a*) a 37½-sec. delay (in work periods the task could be, and was successfully, performed by all *S*s on all trials); (*b*) a further 7½-sec. delay on Trials 1, 4, and 5, or a 5-sec. signal occurring 7½ sec. before puff onset on Trials 2, 3, and 6; (*c*) a 15-sec. puff. The third stage differed from the first two only inasmuch as *S* was allowed to choose during the first 7½ sec. of each trial whether he wanted to work for the puff. It will be noted that, because of the ease of the task and the apparent attractiveness of the puffs, these outcomes were not modified by the performance of the task, since all available puffs were actually received by all *S*s. More importantly, however, the signal allowed no possibility of puff modification by *S* in this situation.

Because the initial instructions relating to the performance of the task were already quite complex, no reference was made at that time to any signal lights. It was assumed that *S*s would notice the signal-puff contingency; when, at the end of the experiment, *S*s were asked to indicate on a printed questionnaire whether they preferred signaled puffs, unsignaled puffs, or had no preference, approximately 80% of the *S*s did not need any explanation of what was meant by "signaled" and "unsignaled." The remaining *S*s, who did ask for an explanation, indicated immediate understanding as soon as *E* mentioned the light, there being no need for *E* to elaborate on the light-puff contingency for any *S*. Finally, *S*s filled out Mehrabian's (1968) Achievement Scale, which differs slightly but appropriately for males and females.

Results and discussion.—Twenty and 3 *S*s, respectively, preferred signaled and unsignaled puffs, while the remaining 17 expressed no preference. This result represents a clearly significant, $p < .01$, preference for information about this unmodifiable but rewarding outcome, although it is worth noting the relatively large proportion of *S*s who expressed indifference about such information. Considering again the possibility that many *S*s were unaware of the signal-puff contingency, it may be suggested that these "indifferent" *S*s were really "ignorant." However, it will be noted that such ignorance would work against obtaining a significant preference for signaled puff, a preference which did clearly emerge in the 40 tested *S*s. Moreover, the proportion of indifferent *S*s approximates those obtained by analogous postexperimental questionnaires from several much simpler aversive conditioning experiments (Furedy & Doob, 1972), where *S*s were clearly aware of signal-shock contingencies (as indicated, for example, by the presence of differential autonomic conditioning), and where no reliable preference for signaled shock emerged in any experiments.

The design of the present experiment also permitted examination of whether the PSP varied as a function of sex, willingness to work for the puff, and need for achievement as measured by Mehrabian's (1968) scales. Chi-square values were calculated for the relationship between the PSP phenomenon and each of the other variables, and the results are summarized in Table 1. It can be seen that the PSP phenomenon is more closely related to sex than to any other variable, with the phenomenon being more marked in females. However, even this sex-based coefficient did not approach significance, so that the results, taken together, do not suggest that the PSP phenomenon is reliably influenced by any of the three variables examined.

TABLE 1

Preference for Signaled Puff as Related to Sex, Work, and Scores on Achievement Scale

Preference	Sex		Work		Achievement scale score	
	Male	Female	High	Low	High	Low
Signaled	8	12	14	6	12	8
Unsignaled	3	0	2	1	2	1
No preference	9	8	9	8	8	9

Note.—χ^2 values (for $p < .05$, $\chi^2 > 7.8$) for sex = 3.86, for work = 1.16, for achievement scale score = .90.

High–Low classifications on the work and achievement variables are based on median splits.

The clear emergence of the PSP phenomenon in the present study contrasts with the repeated failures to find the PSP phenomenon in a number of noxious-outcome studies (Furedy & Doob, 1972). In animal *S*s, a similar contrast can be noted between the noxious-outcome PSS findings quoted by Perkins (1968), which, as mentioned at the outset, did not clearly rule out the possibility of modification, and in the series of carefully designed experiments by Bower, McLean, and Meacham (1966) demonstrating preference for information about rewarding (food) outcomes, even when such information clearly did not permit the *S*s to modify the outcome. In humans, Wiegal and Rodwan (1970) have reported that signaled rewarding outcomes were more attractive than the same unsignaled outcomes. However, although these investigators ruled out the possibility of any modification or "instrumental value" inherent in the signal, the following considerations suggest that signal-produced modification may have played just such a role: (*a*) The outcomes in question were pictures of highly erotically rated nude females presented to male undergraduates, with *S* being both relatively unrestrained, and enjoying a degree of privacy, since he was separated from *E*; (*b*) the greatest signaled–unsignaled attractiveness differences were obtained at signal-outcome intervals of 10 rather than 3 or 1 sec. In short, there is the possibility here of a rather specific form of signal-connected outcome modification which would have instrumental value and which, at least by some of today's undergraduates, is no longer considered to cause blindness.

The discrepancy between unmodifiable noxious and unmodifiable rewarding outcomes as regards the preference for information could be due to the possibility that, in the case of noxious outcomes, the signal elicits conditional fear (e.g., Brown, Kalish, & Farber, 1951), the stimulus consequences of which act as a secondary drive to increase the impact or aversiveness of the noxious outcome (e.g., shock) which follows the signal. Hence, with noxious outcomes, the beneficial effect of the signal through its eliciting, as Perkins (1968) suggests, "preparatory responses" (PRs), may be masked by the detrimental effect of the fear factor, a factor which is not present with signaled rewarding outcomes. It is possible, therefore, to interpret the present PSP effect as supportive of the PR notion of Perkins (1968) as regards rewarding outcomes. However, this confirmation is, at best, indirect. Direct confirmation requires a demonstration that specific forms of the signal-elicited PR do, in fact, increase the attractiveness of the puff outcome in a classical appetitive conditioning situation (e.g., Furedy, 1967). If such confirmation were obtained, PR theory would be as directly supported for rewarding outcomes as the theory has recently been weakened for noxious shock outcomes (Furedy, 1970).

REFERENCES

Biederman, G. B., & Furedy, J. J. The preference-for-signalled-shock phenomenon: Signalling shock is reinforcing only if shock is modifiable. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 681–685.

Bower, G., McLean, J., & Meacham, J. Value of knowing when reinforcement is due. *Journal of Comparative and Physiological Psychology*, 1966, **62**, 184–192.

Brown, J. S., Kalish, H. I., & Farber, I. E. Conditioned fear as revealed by the magnitude of startle response to an auditory stimulus. *Journal of Experimental Psychology*, 1951, **41**, 317–328.

Furedy, J. J. Classical appetitive conditioning of the GSR with cool air as UCS, and the roles of UCS onset and offset as reinforcers of the CR. *Journal of Experimental Psychology*, 1967, **75**, 77–80.

Furedy, J. J. Test of the preparatory-adaptive-response interpretation of aversive classical autonomic conditioning. *Journal of Experimental Psychology*, 1970, **84**, 301–307.

Furedy, J. J., & Doob, A. N. Signalling unmodifiable shocks: Limits on human informational cognitive control. *Journal of Personality and Social Psychology*, 1972, **2**, 111–115.

Lanzetta, J. T., & Driscoll, J. M. Preference for information about an uncertain but unavoidable outcome. *Journal of Personality and Social Psychology*, 1966, **3**, 96–102.

Mehrabian, A. Male and female scales of the tendency to achieve. *Educational and Psychological Measurement*, 1968, **28**, 495–500.

Paul, G. L., Eriksen, C. W., & Humphreys, L. G. Use of temperature stress with cool air reinforcement for human operant conditioning. *Journal of Experimental Psychology*, 1962, **64**, 329–335.

Perkins, C. C., Jr. An analysis of the concept of reinforcement. *Psychological Review*, 1968, **75**, 155–172.

Wiegal, D. B., & Rodwan, A. S. A test of the preparatory response theory by measurement of increased stimulus attractiveness following a signal. *Journal of Experimental Psychology*, 1970, **86**, 225–229.

(Received November 1, 1971

Journal of Experimental Psychology
1972, Vol. 95, No. 2, 472–474

ABSOLUTE JUDGMENTS OF RECENCY FOR PICTURES AND NOUNS AFTER VARIOUS NUMBERS OF INTERVENING ITEMS[1]

JAMES L. FOZARD[2] AND JANE R. WEINERT

Veterans Administration Outpatient Clinic, Boston, Massachusetts

At their own pace, *S*s viewed a filmed inspection sequence of unrelated pictures and nouns. In questions interspersed among the inspection items, *S*s estimated the number of items that separated the presentation of a picture or a noun in a question from its earlier presentation in the inspection sequence. When the number of such items was relatively small (13 or 14), pictures were judged to be relatively more recent than nouns separated by an equivalent number of items, but as the separation became larger (49–50 other items), the differences between the judged recency of the two classes of stimuli decreased. The variability of the judged recency of nouns was larger than that of pictures, especially when the separation between question and inspection sequence was small.

In an earlier experiment (Fozard, 1970), *S*s viewed a sequence of inspection items consisting of unrelated nouns and pictures; in questions, pairs of items were presented to *S* who was required to select the member of each pair he believed he had seen more recently. In some questions the items compared were both nouns or both pictures; in others, one was a picture and the other a noun. Discrimination of relative recency was better when both items presented in the question were pictures than when both were nouns, a finding consistent with those of earlier experiments (Fozard & Lapine, 1968; Fozard & Yntema, 1966; Yntema & Trask, 1963). When the more recent item was a picture and the less recent one a noun, discrimination was poorer than when both items were pictures but better than when both items were nouns. When the less recent item was a picture and the more recent one a noun, discrimination was poorer than when both were nouns. Thus discrimination performance was worse in both cases when the two items were from different rather than the same class of stimuli.

The results of the experiment were consistent with a hypothesis that the average apparent recency of nouns is less than that of pictures separated by the same number of items from their question, and that the judged recency of nouns is more variable than that of pictures. Alternative hypotheses, that nouns and pictures differ only in the variability of their recency judgments or that the two classes of stimuli differ only in the average of their judged recencies, were inconsistent with the rank order of performance in the four conditions.

The hypothesis offered for the observed differences between the apparent recency of pictures and nouns did not consider the manner in which varying the number of intervening items between an inspection item and its question might influence judgments of recency. The major purpose of the present experiment, therefore, was to compare the effect, on the judged recency of pictures and nouns, of varying the number of stimuli separating the presentation of an item in the inspection sequence from the question. The hypothesis evaluated was that the central tendency and variability of recency judgments of pictures would change more with separation than they would with nouns. More specifically, pictures were expected to be judged relatively more recent than nouns at small separations than at large ones. Also, assuming that nouns are effectively closer to the end of *S*'s subjective recency scale than pictures (Fozard, 1970, p. 142), the variability of recency judgments at small separations for nouns was expected to be larger than those for pictures; variability of judgments would, however, increase less with greater separation than would be the case for pictures.

A second purpose of the experiment was to relate recency judgments and the ability to correctly identify old and new items to differences in the ages of the *S*s. Although age-related differences in estimating duration of short time intervals have been reported (see Comalli, 1967), few data have been reported for longer intervals.

Method.—Twenty-one males ranging in age from 25 to 73 yr. and differing widely in educational and socioeconomic status (Warner, Meeker, & Eells, 1960) served as *S*s. They were volunteer participants in a longitudinal study of healthy aging and, by various medical criteria, they were above average in physical health (Bell, Rose, & Damon, 1966).

One of the two equivalent test films described by Fozard (1970) was employed in the present study. The test film was presented to each *S* individually. The *S* advanced the film at his own pace by means of a control button that moved the film, one frame at a time, through a 16-mm. projector. The time required to complete the experiment ranged from 50 min. to a maximum duration of 90 min. allowed by *E*.

---

[1] This research was performed as part of the Veterans
Administration Normative Aging Study, B. Bell, Director. It
was supported in part by National Science Foundation Grant
GB 7837, National Institute of Child and Human Development
Grant HD00340, and the Council for Tobacco Research—
U. S. A. We wish to thank G. D. Carr for his suggestions con-
cerning the analyses of the recognition data.

[2] Requests for reprints should be sent to J. L. Fozard, Veterans Administration Outpatient Clinic, 17 Court Street, Boston, Massachusetts 02108. The author is also at Harvard Medical School.

The film began with a series of 14 frames each containing a single inspection item; 7 were colored pictures and 7 were concrete nouns. The *S* was told to try to remember the inspection items he saw. After the first 14 items, a question frame consisting of 2 items side by side was presented. Questions contained either 2 pictures, 2 nouns, or a picture and a noun. After the first question, a sequence of 2 inspection items, one a picture and the other a noun, was presented and followed by another question; the sequence of 2 new items and a question was then repeated throughout the film.

On each question, *S* was instructed to first inspect the left-hand item. If he believed that the item had not been presented before in the film he signified this by marking an "X"; if he believed it had been presented before as a single inspection item, he was to judge how many frames back in the film he believed it had been presented, and to write the number. He repeated this procedure with the item on the right.

The *S* had been instructed that any given item would never appear more than twice in the experiment, once in inspection and once in a question. He was also told that some items would be presented only in a question.

In some questions, both items presented had appeared earlier as inspection items; in others one of the items had been presented earlier, while the other one had not. (There were no questions in which both items presented in the question had not been presented earlier in the experiment because the film had been used originally in an experiment in which a forced-choice procedure was employed.) The *S*, however, was not told of these constraints.

In summary, *S*s were instructed to make one of two responses to each item in a question: a *numerical* judgment of the recency of any item he believed had been presented earlier in the experiment or a nonnumerical judgment, X, for any item he believed had not been presented earlier in the experiment. Considered as recognition judgments, a response of X to an item presented earlier, signified a failure by *S* to recognize an old item, while an X response to a new item was its correct identification. A numerical judgment to a previously presented item indicated a correct recognition of an old item, but a numerical judgment to an item presented for the first time in a question signified a *failure to distinguish* an old from a new item.

Details of the plan of the test film are described elsewhere (Fozard, 1970). The smallest number of frames on the film separating the initial presentation of an inspection item and its subsequent presentation in a question was either 13 or 14 (average = 13.5) frames. The other two average separations were 25.5 and 49.5 frames. The number of questions containing two items of the same class or one of each was balanced within each pair of separations represented in the questions. There was a total of 36 test items for each combination of separation and stimulus class. In addition, there were 36 pictures and 36 nouns presented for the first time in a question.

Results and discussion.—The basic data analyzed were the medians and interquartile ranges (IQR) of each *S*'s judgments and the fraction of identifications of new and old items. The number of judgments by different *S*s varied: 15 of the 21 completed all of the questions; 4 made between 25 and 30 of

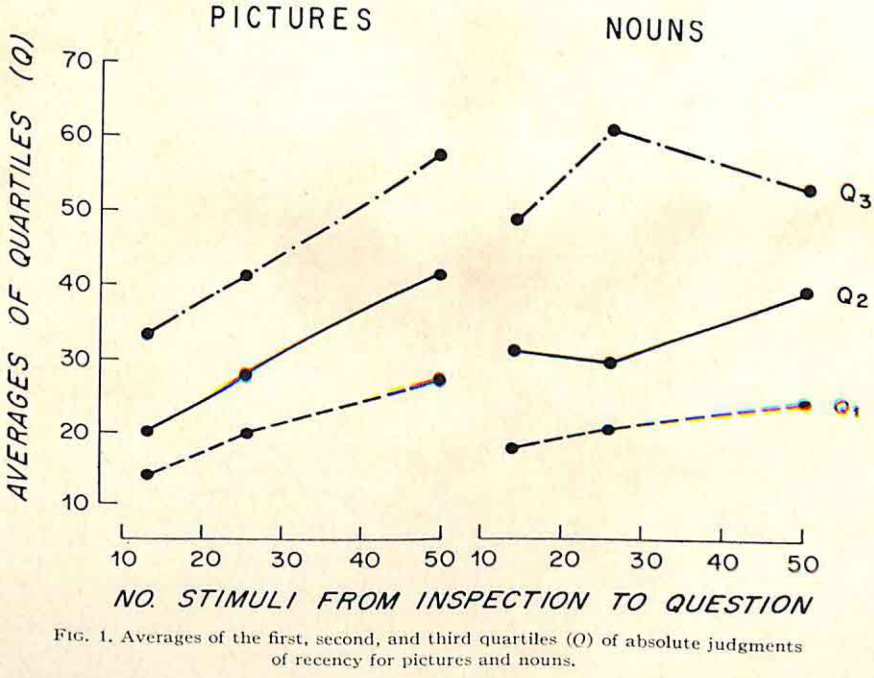


FIG. 1. Averages of the first, second, and third quartiles (*Q*) of absolute judgments of recency for pictures and nouns.

Accept this FREE 144-page guide

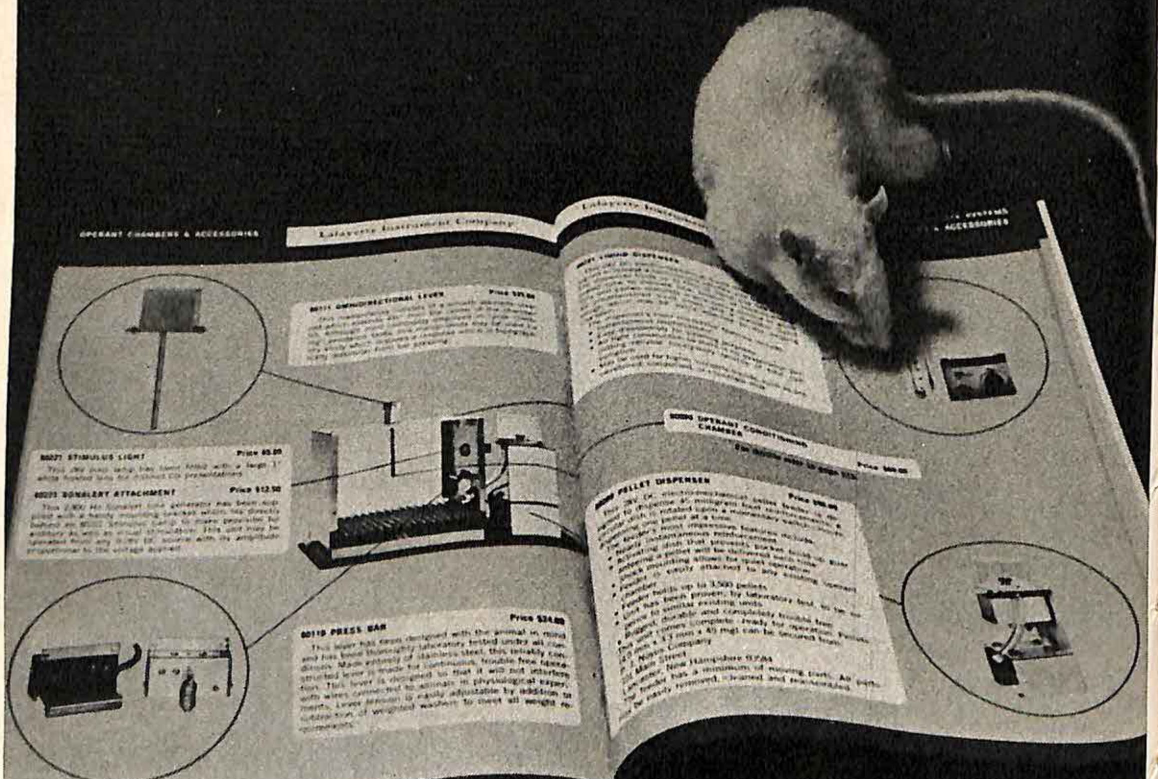
The new 1973 Lafayette Instrument Company guidebook of scientific and laboratory instruments for the life sciences.

On page 106, for instance, you'll see how to set up a custom student operant conditioning chamber. And that's just one page.

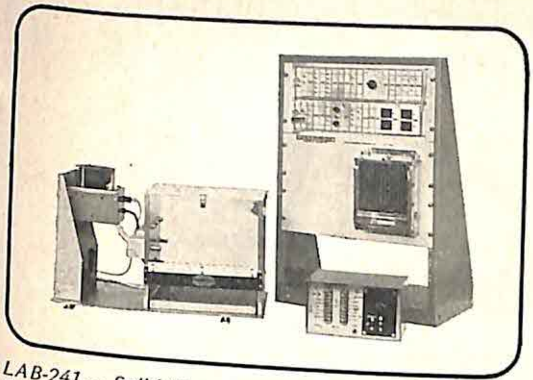
Each page is filled with facts, prices, and information about our newest quality instruments. So naturally, we'd like you to get your hands on our guidebook.

Send for your personal copy today. Ask for the 1973 edition of guidebook no. 1170.

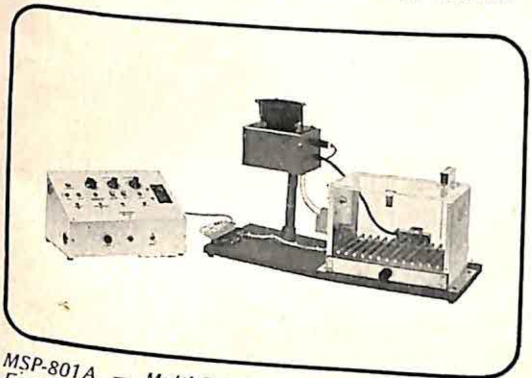
LAFAYETTE INSTRUMENT COMPANY
P.O. Box 1279/Lafayette, Indiana 47902/ 317-423-1505



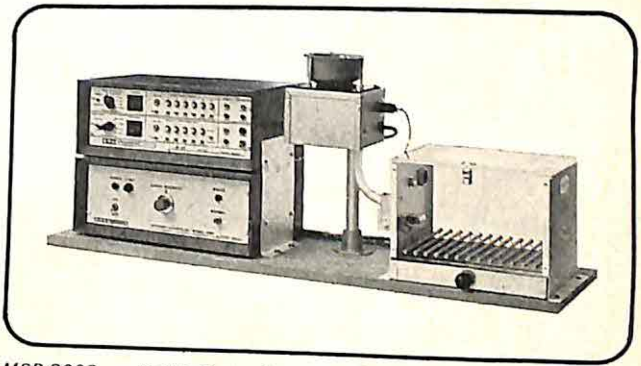
A LAB PACKAGE FOR YOU



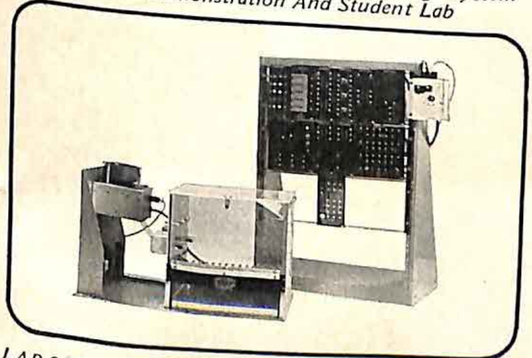
LAB-241 – Solid State System For Student Lab And Research



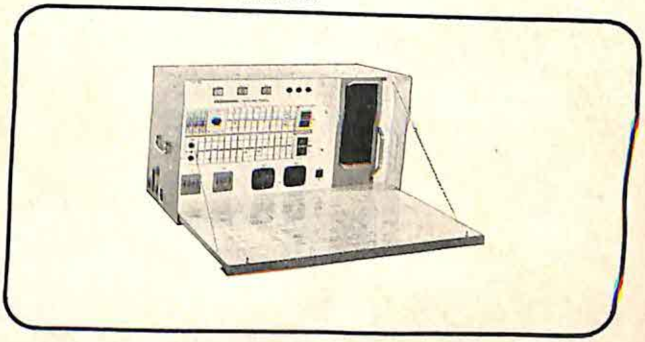
MSP-801A – Multi-Schedule Programming System For Classroom Demonstration And Student Lab



MSP-3003 – Solid State Programming System For Classroom Demonstration And Student Lab



LAB-931 – Electro-Mechanical System For Student Lab And Research



DLC-002 Digi Lab – Self-Contained Digital Logic Console For Student Lab And Research Program Control.

CCL-801 – Classical Conditioning Simulation Unit For Classroom Demonstration And Student Lab

These are just a few of the Lab Packages BRS-Foringer has to offer. One of the more than 60 standard Lab Packages described in Brochure **No. 572** will probably fit your laboratory needs. If you have a special lab system application, BRS-Foringer can take care of that too.

BRS FORINGER®
A DIVISION OF TECH SERV, INC.
5451 Holland Drive Beltsville, Maryland 20705

Call TOLL FREE 800-638-0858*

*Outside Continental U.S. and in the state of Maryland call us collect: 301-474-3400

Iconix makes an electronic switch for audition

Now you can buy a solid-state low-distortion audio switch that:

- ☐ **Switches signal frequencies from 36 kHz down to d-c (i.e. there's no low-frequency cutoff)**
- ☐ **Has a d-c baseline that's easy to balance**
- ☐ **Accepts high- or low-level signals, so it can be inserted at any part of your audio chain**
- ☐ **Has identical rise and fall times with *negligible (low energy) switching transients***
- ☐ **Is quiet when switch is OFF.**

Naturally, the rise and fall times are controllable. So is the channel gain. Price? Only $180.

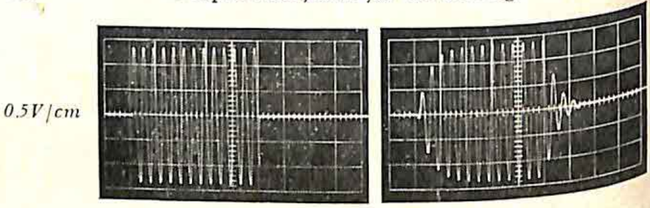
If you like detailed specs, try these.

Input and output Z: 600 ohms
Output level: 1 volt rms into 600 ohms
Gain: variable from 0.2 to 2
Bandwidth: dc to 36 kHz @ 3-dB down
S/N ratio: >60 dB at full signal output
Envelope rise and fall times: 20 microsec. or 1, 2.5, 5, 10, 25, 50, 100 or 250 millisec.

Check these specs against anything else at any price. And if you're doing multi-channel work, you'll find our system prices equally attractive.

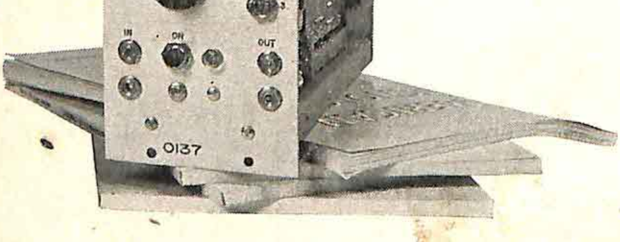
Logic. Timers. T-scopes. The rest of the Iconix line is in our catalog. Write for it. And if an instrumentation problem is puzzling you, drop us a line. We'll work up a solution that gets your experiment under way.

Output waveforms for 2.5 kHz signal.



Rise/fall time: 20 microseconds

Rise/fall time: 1 millisecond



The Iconix 0137 Electronic Audition Switch has minimal switching transients, consistent at all settings.

ICONIX PRODUCTS
Siliconix incorporated
2201 Laurelwood Road
Santa Clara, CA 95054
Hot line for info/service: (408) 243-2424

CONTENTS—(*Continued*)

Interference in Short-Term Motor Memory: Interpolated Task Difficulty, Similarity, or Activity?: B. H. Kantowitz 264
Sentence–Picture Comparison: A Test of Additivity of Processing Time for Feature Matching and Negation Coding: L. E. Krueger 275
Stimulus Modality Effects on Forgetting in Short-Term Memory: D. L. Scarborough 285
Interaural Attention Shifting as Response: L. T. Guzy and S. Axelrod 290
Emotional Images as Mediators in One-Trial Paired-Associate Learning: E. K. Sadalla and S. Loftness 295
Tracking of Serial Patterns: F. Restle and B. L. Burnside 299
Determinants of Induced Amnesia in Short-Term Memory: D. K. Detterman and N. R. Ellis 308
Encoding and Retrieval in Visual Memory Tasks: N. Frost 317
U-Shaped Metacontrast Functions with a Detection Task: S. Cox and W. N. Dember 327
Display Size and the Distribution of Search Times: I. T. Kaplan, W. Metlay, and C. T. Lyons 334
Response Latency as a Function of Hypothesis-Testing Strategies in Concept Identification: R. T. Fink 337
Representation in Deductive Problem Solving: The Matrix: S. H. Schwartz and D. L. Fattaleh 343
Searching for To-Be-Forgotten Material in a Directed Forgetting Task: W. Epstein and L. Wilder 349
The Ponzo Illusion in Stereoscopic Space: R. T. Greene, R. B. Lawson, and C. L. Godek 358
Effects of "Right" and "Wrong" in a Thorndikian Experiment: I.-M. Liu 365
Criterion Effects in Simple Reaction Time: Results with Stimulus Intensity and Duration Manipulations: A. J. Sanford 370
Transfer as a Function of Stimulus, Response, and Simultaneous Stimulus and Response Similarity: B. S. Uehling and B. J. Underwood 375
Facilitation of Human Operant Responding by Stimuli which Precede Aversive Events: R. A. Poetter and P. Lewis 382
Acquisition and Forgetting of Hierarchically Organized Information in Long-Term Memory: T. O. Nelson and E. E. Smith 388
Trade-Off Bias and Efficiency Effects in Serial Choice Reactions: R. G. Swensson .. 397
Conditioned Inhibition of the Rabbit's Nictitating Membrane Response: H. G. Marchant III, F. W. Mis, and J. W. Moore 408
Proactive Interference in Short-Term Recognition and Recall Memory: W. M. Petrusic and R. F. Dillon 412
Free Recall of Grouped Words: R. Gianutsos 419
Auditory Codability and the Short-Term Retention of Visual Information: R. S. Nickerson 429
Temporal Aspects of Simple Multiplication and Comparison: J. M. Parkman 437

Short Reports

Time Estimation and the Interstimulus Interval Function in Classical Conditioning: K. J. Teller, R. Dieter, and M. D. Suboski 445
Vicarious Reinforcement and Model's Behavior in Verbal Learning and Imitation: K. J. Kaplan 448
Escape Maintenance under Serial and Simultaneous Compound Presentations of Separately Established Conditioned Stimuli: D. J. Levis and H. S. Levin 451
The Interrelationship between the Ebbinghaus and Delboeuf Illusions: J. S. Girgus, S. Coren, and M. Agdern 453
Free Recall following a Switch in Encoding Class: M. S. Humphreys, W. M. Petrusic, and R. M. Schwartz 455
"Stroop" Effect: Input or Output Phenomenon?: D. L. Hintzman, F. A. Carre, V. L. Eskridge, A. M. Owens, S. S. Shaff, and M. E. Sparks 458
Associative Reaction Time, Meaningfulness, and Mode of Study in Free Recall: D. Locascio and R. Ley 460
Directionality of Associations in Paired-Associate Learning: S. L. Giurintano 463
Rehearsal of Individual Items in Short-Term Memory: G. F. Meunier, D. Ritz, and J. A. Meunier 465
Effects of High-Priority Events on Recognition of Adjacent Items: L. S. Schulz and R. B. Straub 467
Preference for Information about an Unmodifiable but Rewarding Outcome: J. J. Furedy and F. Klajner 469
Absolute Judgments of Recency for Pictures and Nouns after Various Numbers of Intervening Items: J. L. Fozard and J. R. Weinert 472

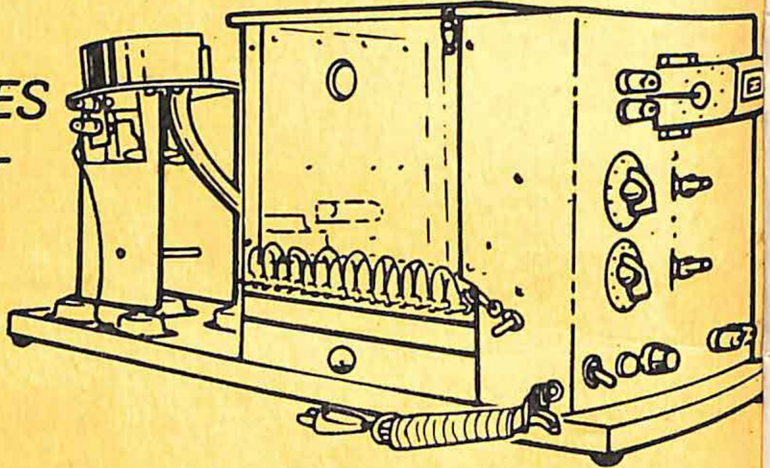
CAGED IN?

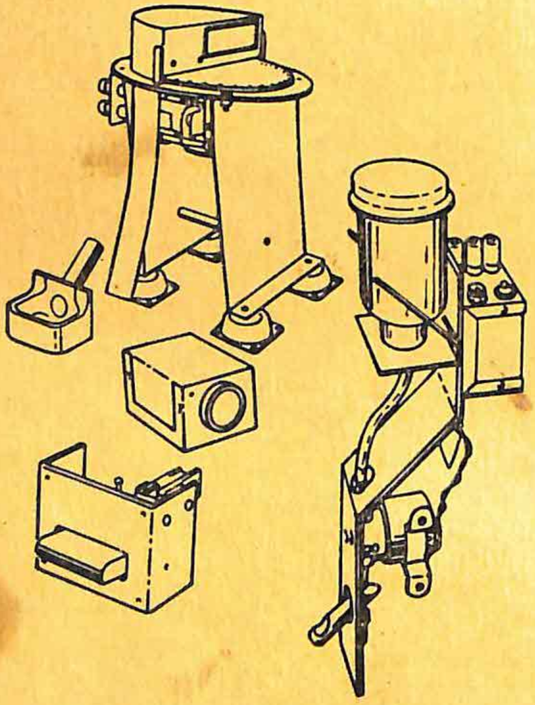
2 FEB 1974

Caged in by inflexible animal testing hardware? Tired of buying cage after cage after cage? Because the cage you bought two months ago, which was fine for its initial applications, cannot be converted to your new research demands?

S.P. INTRODUCES THE ELASTIC CAGE

AND ACCESSORIES AFTER THE FACT





. . . Test Cages designed for flexibility. Units easily converted to new research demands. As they arise.

For example, all A-100 Rodent Cages offer a grid-bar floor already insulated for shock. You can buy a cage for positive reinforcement initially and easily convert it to negative reinforcement schedules. When you want to.

For example, all A-100 Cages have fully interchangeable equipment bearing end panels. So you can change accessories. In minutes.

For example, our accessories.

For examples,
Write for our 1970 catalog.
It's full of examples. Beautiful examples.

Scientific Prototype, of course.

SCIENTIFIC
PROTOTYPE
615 W. 131 ST., NEW YORK, N. Y. 10027, (212) 368-6000